ANALYSIS OF LAYERED SOCIAL NETWORKS

DISSERTATION

Jonathan T. Hamill, Major, USAF

AFIT/DS/ENS/06-03

DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

Research sponsored jointly by the Air Force Research Laboratory, Human Effectiveness Directorate and the National Air and Space Intelligence Center. The United States Government is authorized to reproduce and distribute reprints notwithstanding any copyright notation thereon. The views and conclusions contained in this dissertation are those of the author and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory, the National Air and Space Intelligence Center, the Department of Defense, or the United States Government.

AFIT/DS/ENS/06-03

ANALYSIS OF LAYERED SOCIAL NETWORKS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Jonathan T. Hamill, BS, MS

Major, USAF

September 2006

AFIT/DS/ENS/06-03

ANALYSIS OF LAYERED SOCIAL NETWORKS

Jonathan T. Hamill, BS, MS

Major, USAF

Approved: Date

15 Sept 06

Richard F. Deckro (Chairman)

9-15-06

Alan V. Lair (Dean's Representative)

15 Sep 06

James W. Chrissis (Member)

15 Sep 06

Robert F. Mills (Member)

Accepted:

3 Oct 06

Marlin U. Thomas Date

Dean, Graduate School of Engineering and Management

AFIT/DS/ENS/06-03

Abstract

It is a premise of this research that prevention of near-term terrorist attacks requires an understanding of current terrorist organizations to include their composition, the actors involved, and how they operate to achieve their objectives. To aid in this understanding, operations research, sociological, and behavioral theory relevant to the study of social networks are applied, thereby providing theoretical foundations for new and useful methodologies to analyze non-cooperative organizations. Such organizations are defined as those trying to hide their structures or are unwilling to provide information regarding their operations; examples include criminal networks, secret societies, and, most importantly, clandestine terrorist organizations.

Techniques leveraging information regarding multiple dimensions of interpersonal relationships, inferring from them the strengths of interpersonal ties, are explored. Hence, a layered network construct is offered that provides new analytic opportunities and insights generally unaccounted for in traditional social network analysis. These offer decision makers improved courses of action designed to impute influence upon an adversarial network, thereby achieving a desired influence, perception, or outcome to one or more actors within the target network. In addition, this knowledge can also be used to identify key individuals, relationships, and organizational practices. Subsequently, such analysis may lead to the identification of weaknesses that can be exploited in an endeavor to either eliminate the network as a whole, cause it to become operationally ineffective, or influence it to directly or indirectly support National Security Strategy.

In today's world, proficiency in this aspect of warfare is a necessary condition to ensure United States National Security, as well as to promote and maintain global stability. Quantitative methods serving as the basis for, and discriminator between, courses of action seeking a path towards peace are a principal output of this research.

Acknowledgements

If it weren't for my Dean's rep, I wouldn't have graduated anytime soon.

If it weren't for my committee, I wouldn't have graduated.

If it weren't for Dr. Deckro, I wouldn't have been inspired.

If it weren't for my "red light neighbors," I wouldn't have had great friends.

If it weren't for my loving wife, I wouldn't have anything.

If it weren't for my parents, I wouldn't be here.

Thanks to all.

Go Air Force!

Jonathan T. Hamill

12 Dec 2006

Table of Contents

Page

Page

List of Figures

Figure		Page

List of Tables

List of Abbreviations

Abbreviation	
AFDD	Air Force Doctrine Document
ALN	National Liberation Army
ATO	Air Tasking Order
BDA	Battle Damage Assessment
CA	Combat Assessment
DHS	Department of Homeland Security
DM	Decision Maker
DOD	Department of Defense
DOS	Department of State
FLN	National Liberation Front
GFP	Generalized Network Flow Problem
JI	Jemaah Islamiya
JP	Joint Publication
JTC	Joint Targeting Cycle
KPP	Key Player Problem
MCDM	Multi-Criteria Decision Making
MP	Mathematical Programming
MST	Minimum Spanning Tree
NSS	National Security Strategy
OR	Operations Research
RBAP	Reach-Based Assessment of Position
SDVF	Single-Dimensional Value Function
SMART	Simple Multi-Attribute Rating Technique
SMARTS	Simple Multi-Attribute Rating Technique using Swings
SMARTER	Simple Multi-Attribute Rating Technique Exploiting Ranks
SN	Social Network
SNA	Social Network Analysis
TCT	Time Critical Targeting
TST	Time Critical Target
VFT	Value Focused Thinking

ANALYSIS OF LAYERED SOCIAL NETWORKS

I. Introduction

"To know them means to eliminate them" - Colonel Mathieu in the movie, *Battle of Algiers* [Pontecorvo, 1967].

1.1 Background

This opening quote refers to Colonel Mathieu's objective of quelling the violent insurrection lead by the National Liberation Army (ALN), under direction by the National Liberation Front (FLN). The FLN's struggle for independence from French rule in Algeria relied upon underground organizations not unlike the terrorist networks highlighted in today's media. Colonel Mathieu realized that defeating this elusive organization could only be accomplished by understanding the organization's objectives and its underlying social structure, consequently placing a greater reliance upon intelligence and analysis than mere application of military force.

Truly *knowing* these clandestine organizations means understanding how they arrange and build their structures through recruitment, their underlying motivations for violent and seemingly irrational behavior, and their methods of operational control and execution of terrorist activities. Once gained, this knowledge can then be used to identify key individuals, relationships, and organizational practices. Subsequently, such analysis may lead to the identification of weaknesses that can be exploited in an endeavor to either eliminate the network as a whole, cause it to become operationally ineffective, or influence it to directly or indirectly support our own objectives. In today's interconnected world, proficiency in this type of warfare is a necessary condition to ensure U.S. national security, as well as to promote and maintain global stability.

Table 1.1: Means to Achieve U.S. National Security Strategy

1.	Champion aspirations for human dignity.
2.	Strengthen alliances to defeat global terrorism and work to prevent attacks against us and our friends.
3.	Work with others to defuse regional conflicts.
4.	Prevent our enemies from threatening us, our allies, and our friends with weapons of mass destruction.
5.	Ignite a new era of global economic growth through free markets and free trade.
6.	Expand the circle of development by opening societies and building the infrastructure of democracy.
7.	Develop agendas for cooperative action with the other main centers of global power.
8.	Transform America's National Security Institutions to meet the challenges and opportunities of the twenty-first century.
9.	Engage the opportunities and confront the challenges of globalization.

The National Security Strategy (NSS) focuses not rather on the "great armies" and countries with "great industrial capabilities" as was required in the past, but upon the "shadowy networks of individuals [that] can bring great chaos and suffering to our shores for less than it costs to purchase a single tank" [The President, 2006, np]. The second of eight means (see Table 1.1) to achieve the NSS–*Strengthen alliances to defeat global terrorism and work to prevent attacks against us and our friends*–deals directly with the subject of terrorism and how to counter its effects upon the nation. Specifically, it necessitates the disruption and destruction of global terrorist networks via attacks upon their "…leadership; command, control, and communications; material support; and finances" [The President, 2006, pg. 1].

It can also easily be argued that the remaining means, at a minimum, indirectly focus upon conquering terrorism, either by reducing the conditions that give rise to future terrorists or by minimizing the risks associated with our country's vulnerabilities to terrorist acts. These latter efforts are led by the Department of Homeland Security (DHS). Accordingly, the prioritized strategic objectives of DHS include: "prevent terrorist attacks within the US; reduce America's vulnerability to

terrorism; and minimize the damage and recover from attacks that do occur" [DHS, 2002, pg. vii].

It is a premise of this research that prevention of near-term terrorist attacks requires an understanding of current terrorist organizations to include their composition, the actors involved, and how they operate to achieve their objectives. Further, the prevention of far-term terrorist attacks requires a number of preemptive measures both external and internal to current terrorist organizations. Externally to the existing terrorist networks, the economic, social, and political conditions that contribute to the recruitment process must be addressed; this appears to be the upcoming, preferred U.S. strategy. Internally, current members must be convinced (one way or another) to discontinue the use of violent behavior as the primary means of achieving their political objectives. The mid-term realm is hypothesized to consist of a continuous struggle of a combination of offensive (counterterrorism) and preventive (anti-terrorism) measures that facilitate the transition over time to a long-term approach. The realm of primary interest in this research is to support near-term, counterterrorism efforts.

1.2 Problem Definition

The overarching objective of this research is to expand operations research, sociological, and behavioral theory relevant to the study of social networks, thereby providing theoretical foundations for new and useful methodologies to analyze non-cooperative organizations. Social networks are classically defined as "the set of actors [individuals] and the ties [relationships] among them" [Wasserman and Faust, 1994, pg. 9]. For the purposes of this research, non-cooperative organizations are those trying to hide their structures or are unwilling to provide information regarding their operations; examples include criminal networks, secret societies, and, most importantly, clandestine terrorist organizations [cf., Sparrow, 1991; van Meter, 2002].

Given the resultant understanding and insights provided by the analytic techniques developed in the course of this research, decision makers are offered better courses of action that impute influence upon the adversary's network. Such courses of action seek to achieve a target influence, perception, or outcome to one or more actors within the network through either direct or indirect means. The method or methods of imposing influence upon a network may take on a number of forms; for example, specific individuals may be directly targeted or a number of relationships between two or more individuals may be exploited or altered via information, influence, or psychological operations. Further, kinetic, non-kinetic, or a mix of approaches, may be considered to achieve these methods of influence. In addition, the time-line upon which influence is applied could range from immediate to long-term; consequently, options to mitigate near- and far-term terrorist activities can be explored, respectively.

Quantitative methods serving as the basis for, and discriminator between, courses of action are a principal output of this research. The quantitative study of social networks has been undertaken in 'modern' sociological and anthropological studies for some time, beginning with a graphical representation of the social network known as the *sociogram* [Moreno, 1953]. However, the majority of applications have been primarily descriptive in nature, focusing on "[measuring] interpersonal relations in small groups ..., [describing] properties of social structures and individual social environments ..., and [assessing] the impact of structural arrangement on group problem solving and individual performance..." [Wasserman and Faust, 1994, pg. 4, 12-13]. There exist a number of works that have investigated clandestine organizations, which are included within the context of 'non-cooperative,' networks from a SNA perspective [e.g., Sparrow, 1991; Carley et al., 2002; Krebs, 2002; van Meter, 2002; Sageman, 2004; Xu and Hsinchun, 2004]. The research presented here builds upon such sociologically-oriented methodologies as well as the recent operations research-oriented works of Renfro [2001], Sterling [2004], and Clark [2005],

all of which have bridged gaps between descriptive sociological and anthropological techniques and prescriptive operations research techniques.

As Renfro and numerous other works within the genre of SNA suggest, the commodity of influence flows through these social networks, a phenomenon often used to study the spread of rumors, acceptance of innovations, coalescence of group opinions, and so forth. However, the rates and capacities of the conceptual flow of interpersonal influence are predicated upon the situation, the organizational norms, the relationships between, and the individual characteristics held among the interacting individuals. Analyses that often yield asymmetry of relationship existence between actors include 'who works for whom,' which individuals are opinion leaders or early adopters of innovation, or even 'who is friends with whom' where some respondents forget friends [cf., Brewer and Webster, 1999]. Alternatively, asymmetry of influence over other individuals may be conjured by the classic 'E. F. Hutton' example or any other leader-follower type of relationship[1]. Further, the direction of dominance may change, given a change in context. For example, the team captain may not be the class leader. Means to measure and take advantage of these effects of influence within operations research methods provide an opportunity to improve the current social network modeling capabilities, particularly those having underlying assumptions not amenable to the study of non-cooperative networks.

Renfro also discussed multiple contexts of relationships between individuals and posited an accompanying multi-commodity network formulation; however, there remain numerous research questions and clarifications of these concepts and their underlying theory. For example, when people interact, there is often more than one affiliation or context upon which that interaction may benefit (e.g., past friendships, familial, education, religious, and professional contexts). Consequently, modeling

[1]Recall the advertisement for the brokerage, E. F. Hutton. The typical scene involved a crowded, noisy, room. When a representative of this firm began to speak, all other conversation immediately stopped in order to hear what advice or information was being offered.

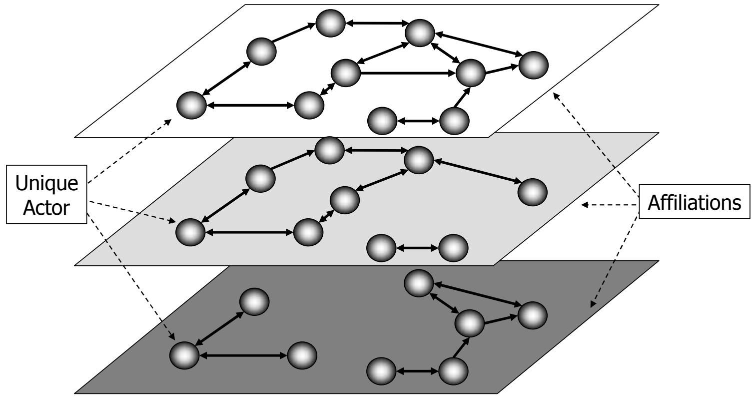


Figure 1.1: Layered Social Network

these contexts as layers of affiliation (see Figure 1.1) may provide a means to gain new insight into how, why, and to what extent dyadic interactions occur.

1.3 General Assumptions and Scope

Admittedly, social interaction, and therefore the subsequent network composition and/or topology, is far from a static process. As time progresses, friendships can strengthen, weaken, or even disappear altogether. Additionally, new friendships may be formed, either among individuals already within the network, incorporating entirely new members, or both. Finally, current members may leave the network over time due to any number of reasons (e.g., alienation by the group, renunciation of one's membership to the group, capture, or death). Consequently, a social network is a dynamic entity.

Interestingly, most studies investigating the propagation of opinions, rumors, or influence over time throughout a given social network assume a known, and unchanging, network topology. This is clearly contrary to social interaction observed in the real world, even for non-cooperative networks that minimize their social interaction in an effort to maintain secrecy. Structural change over time is inevitable due to the risky endeavors often undertaken (i.e., actors and their associated links disap-

pear due to capture or death); further, the never-ending need to ensure survival of the organization and the fight for its cause necessitates the creation of new bonds of trust (i.e., recruitment, shared experience, and so forth). Current developments attempting to overcome this limiting assumption include the computational modeling, essentially an agent-based simulation approach proposed by [Monge and Contractor, 2003, Chapter 4], and the dynamic network analyses techniques proposed by [Carley, 2003; Marsili et al., 2004].

Unsurprisingly, dynamic analysis is almost certainly more representative of real-world social network systems and is perhaps the ultimate modeling goal, particularly for the analysis of long-term anti-terrorism approaches. However, the dynamic approach subsumes a range of data requirements. For example, in order to model actor behaviors, one must know, or hypothesize at a minimum, the possible courses of action available to the actor conditioned upon a multitude of environments, the likelihoods of the individual taking one of those actions, the subsequent behavior of other actors in reaction, and so forth.

Another potential issue for viewing and analyzing this problem dynamically is the underlying assumption of the time-sensitive nature of counterterrorist operations. It is assumed that, in general, the time-sensitive aspect will be weighed against both data uncertainty and availability. Said another way, this type of analysis will likely not enjoy the general assumption of accurate and complete data that prevails throughout the SNA literature [Thomason et al., 2004].

Uncertainty in the data of both individual and overall network characteristics will remain an underlying concern despite continuing investigative efforts. Related dilemmas have been previously studied by Stork and Richards [1992] regarding nonrespondents in studies (akin to a terrorist not answering the question “who are your fellow terrorists?”); the effects of unintended (or intended, depending upon the context) asymmetries upon structural properties of friendship networks [Brewer and Webster, 1999]; and the impact upon classical SNA centrality measures as a result

of incorrect node or edge data Borgatti et al. [2006]. Over time and with more investigative resources, reliability and confidence in the data may increase; however, the amount of time reserved for intelligence gathering to reduce this uncertainty is also provided to the terrorists and their recruiting, operations planning, and execution activities.

Given these considerations, methodologies that can be implemented quickly to accommodate updated intelligence and are robust enough in the face of uncertainty to still permit useful quantification of network phenomena are desired. Sensitivity and parametric analysis enable testing and characterization of the requisite assumptions and are assumed to sufficiently alleviate data concerns. Consequently, this research assumes that a static picture of the network at given points in time is sufficient for the methodologies proposed for two main reasons: scarcity and questionable accuracy of available data and the potentially time-sensitive nature of operations that may apply the methodologies herein.

As in the real-world, research data characterizing not only non-cooperative networks but the contexts within which they interact is difficult to obtain. In order to study these multi-level networks and test such measures, data composed of either open, unclassified sources or of generated notional networks are presented.

As previously noted, short-term goals for the global war on terror include disbanding, disrupting, and eliminating current organizations; long-term goals seek to reduce the underlying conditions favorable to terrorism. Although the long-term strategy is likely the only way to ensure our nations's future security, the short-term strategy will not only improve our immediate security but will contribute to the long-term strategy as well. Therefore, this research focuses primarily upon short-term strategies that implement counterterrorist options.

In order to develop these options, new measures of inter-personal influence and modeling techniques to employ them are proposed, accommodating when possible the nuances of non-cooperative networks. These measures and their accompany-

ing theory are compared to existing ones when possible for verification purposes. Additionally, previous mappings between SNA and OR, specifically in the area of network flow models, require further work in the area of flow typology. Overall, this research lends itself to improved modeling capabilities regarding the impact of influence operations upon the individuals within a non-cooperative network.

1.4 Research Objectives

The primary objective of this research is to develop the underlying theory and associated methodology used to generate and analyze courses of action that may be applied to networks of non-cooperative individuals. The courses of action specifically seek to shape the intentions of our adversaries through influence. This activity is within the context of military psychological operations that strive to influence an adversarys "…emotions, motives, objective reasoning, and ultimately, the behavior of their governments, organizations, groups, and individuals" [DOD, 2006a, pg. II-1]. These activities are often undertaken in order to achieve a given political goal. The ability to quantitatively assess potential courses of action both prior to execution–to facilitate decision making and alternative selection–and after–to determine the operation's efficacy–serves as a crucial first step to improve our understanding and execution of warfare that extends beyond the realm of physical damage.

To accomplish these objectives, previous SNA and OR cross-sectional works are built upon. This research differs from earlier works in methodological scope as well as the investigation of benefits, disadvantages, and incorporation of layered network perspectives and information into various mathematical techniques. Specific objectives of this research include:

1. Develop a new centrality-like measure, via extensions of several others in use, to screen networks for potential actors of interest. The theoretical bases that

make this measure more amenable to non-cooperative networks, advantages, and computational challenges are presented.

2. Develop new techniques to identify key members of an organization in line with the 'key-player problem' described by Borgatti [2003a]. Mathematical programs equivalent to current heuristic approaches are presented and compared. Further extensions of the programs, and therefore the technique, are developed and discussed to accommodate specific analysis requirements as well as other methodological constructs presented in other areas of this research.

3. Develop a theory of measuring interpersonal relationships accounting for multiplexity. This measurement approach facilitates the incorporation of multiplexity into mathematical programs of social networks. Methods to characterize and analyze non-cooperative networks as layered, inter-dependent networks are also investigated. In addition, determining which contexts or layers are of interest, given a specified organization and scenario, is explored.

4. Exploration and explanation of how to aggregate, as appropriate, multiple social networks into a single, weighted network upon which classical and newly developed analyses may be performed. In addition, if psychological operations are applied to one or more layers, investigation of how the weights may change over time and the affect upon the network performance and exchange of influence (or power, or status, etc.) measures in response to these external forces (courses of action) are performed.

5. Extend current power theory and develop a pair-wise valued measure of gains, losses, or thresholds of influence between two individuals. Incorporation of this measure into generalized network flow formulations and its subsequent impact upon analysis methodology and results are explored. A new measure, an extension of the centrality measure proposed by Freeman et al. [1991], is developed.

6. Apply the mathematical modeling, decision-analysis-like techniques, newly developed social network measures, and other, related theory to several, unclassified examples–presenting a process that provides actionable information facilitating course of action development, analysis, and improved capability to forecast roles and responsibilities of individuals in a non-cooperative network when faced with limited information.
7. Combine the most promising techniques into a prototype tool-set, developed in MATLAB, for intelligence analysis use by the sponsoring organizations.

1.5 Dissertation Overview

The organization of this dissertation document is as follows. Chapter II presents the literature relevant to the problem areas and builds the case for the contribution objectives described above. Chapter III provides an overview of the complete methodology and its general assumptions. Chapter IV presents a new social network measure that addresses some limitations of currently available approaches which can be used as a screening technique if limited information exists. Chapter V presents a mathematical programming formulation for a concept amenable to targeting key individuals within a social network. Chapter VI develops and discusses various means to measure the multiplexity of a relationship, which serves as a proxy for the strength of an interpersonal relationship. Chapter VII explores the nuances of persuasion and power theory in order to estimate gains and losses of information or influence as a function of sender-receiver interactions. Although smaller examples are provided for illustrative purposes throughout the document, Chapter VIII details a larger example, where all aspects of the research methodology are applied and discussed. Chapter IX provides overall, general conclusions as well as recommendations for future research.

II. Literature Review

This chapter serves several purposes. First, a brief review of graph theoretic definitions provide underlying terminology. Second, a review of the history and recent developments of social network analysis (SNA) literature is provided, with the latter focusing on the investigation of clandestine (e.g., criminal or terrorist) organizations. The behavioral literature and current modeling efforts of such networks is summarized, providing a basis of understanding for the extremist phenomenon.

In light of the ultimate goal of negating the threat of terrorist networks via the application of influence, ranging from psychological operations to lethal force, underlying theories related to influence and motivation are presented. A number of these theories provide an opportunity to quantify interpersonal influence, albeit some require various assumptions that may or may not lend themselves to the analysis of non-cooperative networks. Opportunities to merge these social network techniques with those of operations research (OR), similarly accomplished by Renfro [2001], are explored.

Several operations research methods have been developed to measure interpersonal influence, incorporating this information into various models to study organizational phenomena [Renfro, 2001; Clark, 2005]. Possible areas of opportunity to enhance existing theory, as well as potential improvements to accommodate non-cooperative network phenomena, are suggested. Obtaining information characterizing non-cooperative networks is fundamentally challenging. Consequently, this research suggests that viewing interpersonal relationships as multidimensional, as opposed to the generally single dimensional assessment, offers a means to improve upon existing models of social networks. Measurement of the strength of interpersonal relationships is derived from a combined and weighted assessment across multiple relationship contexts or dimensions. This research area is derived from de-

cision analytic techniques and its underlying theory of attribute weighting, as well as appropriate graph theory and SNA-related literature.

Lastly, leveraging the output of these techniques as input to a variety of mathematical programming techniques comprise a major focus area of this research. As such, applicable mathematical programming models are reviewed, as well as existing sociological techniques that may benefit from such techniques. As noted by Borgatti [2005], the underlying assumptions of information and influence flow play pivotal roles when choosing a specific social network measure and interpreting its output. Consequently, mismatches between the type of flow and the type of measure applied often results in erroneous conclusions. The typology of flow processes within a network and the implications upon network flow formulations are examined. Overall, the combination of techniques derived or extended from these fields comprise the various elements of the research methodology.

2.1 Graph Theory

While not a complete review of graph theory, this section serves to establish a common graph-theoretic lexicon applicable to the SNA methods of interest within the remainder of this dissertation. Social networks are typically represented and analyzed via a graph [Moreno, 1953]. A graph, $G = (N, A)$, is comprised of the set N of n nodes corresponding to the individuals, and the set A of m arcs representing the ties, relationships, bonds, or some other contextually-dependent connection between two individuals [Newman, 2003, pg. 173]. An arc going from actor i to actor j is denoted (i, j). Such relationships can be undirected or directed, the latter resulting in a *digraph*. Undirected relationships are symmetric, or $(i, j) = (j, i)$, implying that the relationship, bond, and so forth runs equivalently in either direction. If the context of the sociometric data (e.g., accounting for supervisory roles) or if responses within a sociological survey are not equitable between two given actors

(e.g., the forgetting of friends), a directed, asymmetric arc is more appropriate [cf. Brewer and Webster, 1999].

The number of nodes within a graph denotes the *order* of the graph. The *density* of the graph is the ratio of the number of edges to the theoretical maximum number of edges possible. The maximum number of edges is $n(n-1)/2$ for a graph and $n(n-1)$ for a digraph. Given that the individuals and their existing relationships are often the focus of analyses, a number of SNA measures attempt to describe how information, influence, rumors, adoption, and other influences may flow through the network as a result of its topology.

A *walk* is defined as "a sequence of nodes and lines, starting and ending with nodes, in which each node is incident with the lines following and preceding it in the sequence" [Wasserman and Faust, 1994, pg. 105]. SNA literature generally allows both nodes and arcs to be repeated within a walk [Wasserman and Faust, 1994, pg. 105] whereas graph-theoretic literature does not [Deo, 1974, pg. 19]. Within the set of walks are trails and paths. A *trail* is "a walk in which all of the lines are distinct, though some node(s) may be included more than once" [Wasserman and Faust, 1994, pg. 107]. A *path* is defined as "a walk in which all nodes and all lines are distinct" [Wasserman and Faust, 1994, pg. 105]. Additionally, given a digraph, the term path also implies that the direction of the arcs within the path also follow the direction of the path; otherwise it is a chain [Bazaraa et al., 1990, pg. 422]. The length of a walk, trail, path, or chain is determined simply by the summation of the lengths of each of its arcs. Paths, directed walks with unique arcs and nodes, are taken advantage of within a new social measure, discussed later.

A common assumption underlying many SNA measures, either explicitly or implicitly, requires that information or influence flow along the shortest path within a network–termed a *geodesic path*. The geodesic path is defined as "the [not necessarily unique] shortest path through the network from one vertex to another" [Newman,

2003, pg.173]. The length of the longest geodesic path between all possible node pairs defines the *diameter* (D) of a network [Newman, 2003, pg.173].

Although graph theory considers the nature of all types of networks, this research pays particular attention to previous efforts related to the study of social networks and other types of graphs demonstrating properties similar to those induced by the social interaction of individuals. Next, the concepts of 'random' and 'small world' graphs are briefly discussed and compared to the network that appears to lie between these realms–the social one.

2.2 Graphs and Social Networks

Recent works such as Watts [1999], Barabási [2002], and Buchanan [2002] have popularized what is referred to as the 'small-world' network phenomena. The initial concept of small-world networks is often credited with Milgram's research that investigated the passing of letters to an unknown individual via known contacts. By tracing the paths taken by the correspondence, the famous 'six degrees of separation' between ostensibly distant and unconnected actors was observed [Milgram, 1967].

Such networks are generally contrasted with the random graphs concept developed by Erdos and Renyi [1959] due to the underlying processes that form the links between nodes. However, random and small-world graphs do share some common properties. The use of random networks to study social network phenomena has been attempted, but with mixed results [cf. Newman et al., 2002; Newman, 2003; Borgatti et al., 2006].

2.2.1 Random and Small-World Graphs

In his book, <u>Small Worlds</u>, Watts compares and contrasts the properties of both random and small-world graphs, often referred to within the literature as exponential and scale-free graphs, respectively [Watts, 1999]. Watts details the formal definition

of random graphs, which are essentially "…a vertex set, consisting of n vertices, and an edge set that is generated in some random fashion" [Watts, 1999, pg. 36]. Two prevalent approaches used to develop random graphs are defined as follows.

Definition 1. *$G(n, M)$ is a labeled graph with vertex set $V(G) = \{1, 2, \ldots, n\}$, having M randomly chosen edges (where M usually depends on n). $G(n, M)$ is frequently abbreviated as G_M* [Watts, 1999, pg. 36].

Definition 2. *$G(n, p)$ is a labeled graph with vertex set $V(G) = \{1, 2, \ldots, n\}$, in which every one of the possible C_2^n edges exist with probability $0 < p < 1$, independent of any other edges. $G(n, p)$ is frequently abbreviated as G_p* [Watts, 1999, pg. 36].

Noting that the degree of a vertex (k) is defined as the number of edges incident to that vertex, the distribution of values for degrees among the vertices is one of the differences between random and small-world graphs [Albert et al., 2000, pg. 379]. The degree distribution of vertices within networks developed as either G_M or G_p tend to be homogenous, with exponential-like distributions, implying that vertices with high degree are unlikely [Albert et al., 2000, pg. 379]. On the other hand, the degree distribution of small-world networks is often described as inhomogeneous, following the power law distribution described by $P(k) \sim k^{-c}$ [Albert et al., 2000, pg. 379]. Consequently, highly connected nodes are statistically unlikely in random (or exponential) networks and statistically significant in small-world (or scale-free) networks [Albert et al., 2000, pg. 379].

The term 'scale-free' was coined by Barabási and his colleagues during their investigation of link distribution of the Internet. The power law observation, and therefore the lack of a 'bell-shaped' distribution that was expected, lead to Internet nodes that "defied explanation, almost as if they had stumbled on a significant number of people who were 100 feet tall …" [Barabási and Bonabeau, 2003, pg. 53]. As Buchanan explains,

> [The] power-law distribution is special in that there is no "typical" number of links. In other words, the network has no inherent bias to produce

> elements with an expected number of links; rather this number varies widely over a huge range. That is to say, there is no inherent "scale" for the number of links, and the network is scale-free [Buchanan, 2002, pg. 215].

Numerous connections between real-world, emergent networks and small-world (or scale-free) phenomenon have been made. Examples include cellular metabolism, Hollywood movie stars, Internet connections and world-wide-web page links, protein regulatory networks within cells, research collaborations, and sexual relationships [Buchanan, 2002; Barabási and Bonabeau, 2003, pg. 54].

As Buchanan points out, a purely random network has the small-world property as far as connectivity is concerned [Buchanan, 2002, pg. 54]. Essentially, the random networks exhibit the 'six degrees of separation' effect. However, Buchanan also states that "...for 1,000 people linked together randomly, the degree of clustering turns out to be about 0.01, which is not even close to what one finds in a real social network" [Buchanan, 2002, pg. 54].

Clustering, the main differentiating characteristic between random and small-world graphs, is a topological consequence that results from the underlying "motivations for why we create, maintain, dissolve, and reconstitute our communication networks" [Monge and Contractor, 2003, pg. 223]. Such motivations fall into one of three theoretical mechanisms which include homophily, proximity, and social support [Monge and Contractor, 2003, pg. 223].

Homophily infers that people are more likely to communicate with others similar to themselves, classically referenced by the adage "birds of a feather, flock together" [Monge and Contractor, 2003, pg. 223]. Additionally, Brass reiterates that homophily has "been operationalized on such dimensions as age, sex, education, prestige, social class, tenure, and occupation," and suggested that similarity between two actors is "thought to ease communication, increase predictability of behavior, and foster trust and reciprocity" [Brass, 1995, pg. 51].

Proximity mechanisms assume that closer distances, either physically or, in light of today's computer networked world, electronically, facilitate "...the likelihood of communication by increasing the probability that individuals will meet and interact" [Monge and Contractor, 2003, pg. 227].

Lastly, social support theories focus "...on the ways in which communication networks help organizational members to cope with stress" [Monge and Contractor, 2003, pg. 235]. Relationships, and therefore network connections, are developed as a result of the individual's need to achieve mental well-being, such as the need to belong or the need to discuss personal problems, and determine potential solutions, with an empathetic individual [Monge and Contractor, 2003, pg. 235-6].

In order to more clearly differentiate between random and small-world networks, Watts provides definitions of the network properties *characteristic path length* and *clustering coefficient* that, when combined, serve as the accepted technical and mathematical definition of a small-world network. Such a network is assumed to be connected.

Definition 3. *The characteristic path length* (L) *of a graph is the median of the means of the shortest path lengths connecting each vertex* $v \in V(G)$ *to all other vertices. That is, calculate* $d(v,j)\,\forall j \in V(G)$ *find* $\bar{d}$ *for each* v*. Then define* L *as the median of* $\{\bar{d}\}$ [Watts, 1999, pg. 29].

Definition 4. *The clustering coefficient* (γ) *of a graph characterizes the extent to which vertices adjacent to any vertex* v *are adjacent to each other. Therefore,* $\gamma = 1$ *implies that the graph consists of* $n/(k+1)$ *disconnected, but individually complete subgraphs (cliques), and* $\gamma = 0$ *implies that no neighbor of any vertex* v *is adjacent with any other neighbor of* v [Watts, 1999, pg. 33].

The combination of Definitions 3 and 4 serve as the formal graph-theoretic definition of a small world network.

Definition 5. *A small-world graph is a graph with* n *vertices and average degree* $\bar{k}$ *that exhibits* $L \approx L_{random}\left(n, \bar{k}\right)$*, but* $\gamma \gg \gamma_{random} \approx \bar{k}/n$*, where the asymptotic limit of* L_{random} *is* $\ln(n)/\ln\left(\bar{k}\right)$ [Watts, 1999, pg. 56, 114].

The small-world definition essentially states that this type of network "displays considerable local connectedness while also having a low degree of separation with the other nodes in the network" [Monge and Contractor, 2003]. Interestingly, although a variety of real-world networks, including some social networks, have been shown to adhere to the small-world network construct, it does not necessarily mean that all social networks are small-world networks, particularly when the networks of interest are non-cooperative by nature.

As an example, Krebs applied SNA to the hijacker network that perpetrated the 9-11 attacks. The network, discernable only after the fact and composed of 19 individuals, had a measured characteristic path length of $L_{terrorists} = 4.75$ and a clustering coefficient of $\gamma_{terrorists} = 0.49$ [Krebs, 2002]. The average degree of this network was $\bar{k} = 3.47$. Using this information and Definition 5, a comparable small world network would expect $\gamma_{random} \approx 3.47/19 = 0.183 \ll 0.49$ and $L_{random} = ln(19)/ln(3.47) = 2.36$. The clustering coefficient appears to meet the small-world criteria. However, the path length ($L_{terrorists}$) is longer than expected.

After adding six additional links, based upon what Krebs presumed would be necessary and logical in order for the terrorists to conduct their operations, $L_{terrorists}$ was reduced to 2.79–much closer to the value expected by definition. Krebs concludes that the clandestine nature of this network forced the members to limit their communication and connectivity in order to prevent detection.

This observation is in agreement with the findings of Baker and Faulkner who, in studying organized crime in the electrical industry, determined that the structural development of such networks "...is driven primarily by the need to maximize concealment, rather than the need to maximize [information] efficiency" [Baker and Faulkner, 1993, pg. 837]. Consequently, non-cooperative networks may not necessar-

ily exhibit the small-world properties, either due to our limited knowledge of their organizational structure, the inherent nature of their operations, or both. Fortunately, this neither disqualifies them as social networks nor prevents the application of SNA methods upon them.

2.2.2 Social Network Analysis

The field of social network analysis is often traced back to the work of Moreno [1953], who developed the sociogram, a pictorial representation of a social group via a graph. Thus, with the natural connection to graph theory, Moreno devised a means to quantitatively study the qualitative nature of relationships among individuals within a social grouping [Moreno, 1953]. Subsequently, a variety of tools and techniques have been developed to study the structural nature of social networks and the implications of topology and personal characteristics upon overall network behavior. Most of these techniques perform calculations upon the mathematical representation of the sociogram, the *sociomatrix* ($\mathbf{X}$).

The sociomatrix is a two-way matrix, "indexed by the sending actors (the rows) and the receiving actors (the columns) ...," and is equivalent to the adjacency matrix of a graph when the sociogram captures dichotomous, symmetric relationships [Wasserman and Faust, 1994, pg. 77]. For a given relation ($\Re$) and the set of g actors in $N = \{n_1, n_2, \ldots, n_g\}$, the value $x_{ij} \in \mathbf{X}$ is equal to the value of the tie from n_i to n_j on relation $\Re$ [Wasserman and Faust, 1994, pg. 79-80]. For dichotomous relationships, this value is either 1 or 0 for actors that are or are not adjacent, respectively; however, when x_{ij} is not limited to this data type, it is a valued relation. Although the majority of sociometric studies within the literature focus upon a single relationship during analysis, multiple relationships may be evaluated.

Suppose there exist R relations $\Re_1, \Re_2, \ldots, \Re_R$, each measured on the same set of actors. The value $x_{ijr} \in \mathbf{X_r}$ is the value of the tie from n_i to n_j on relation $\Re_r$. This 'super-sociomatrix' approach offers a means to capture the layers of relations

Table 2.1: Meta-Matrix [Carley et al., 2002, pg. 83]

	Agents	Knowledge	Tasks
Agents	Social Network	Knowledge Network	Assignment Network
Knowledge	–	Information Network	Needs Network
Tasks	–	–	Task-Precedence Network

as depicted in Figure 1.1 [Wasserman and Faust, 1994, pg. 86-7]. Note that the collection of sociomatrices is also a collection of simple graphs, those accounting for only one relation [Wasserman and Faust, 1994, pg. 145]. A multigraph, or multivariate graph in the case of directed relationships, is "a generalization of a simple graph or digraph that allows more than one set of lines" [Wasserman and Faust, 1994, pg. 146]. The method of storage and subsequent approaches to calculate measures using this type of multi-dimensional data is one of the key underlying questions of interest in this research.

Carley et al. [2002] proposed a related concept describing a composite network that incorporates the multi-dimensionality of interpersonal relations is the meta-matrix. The meta-matrix concept is based upon the premise that network dynamics are functions of (1) the social structure, (2) the distribution of knowledge and information, (3) the interrelations between domains of knowledge, and (4) the distribution of work and requirements [Carley et al., 2002, pg. 83]. These network-related aspects of an organization within the meta-matrix construct serves as input into an agent-based network simulation, which evaluates the organization's ability to perform tasks, communicate effectively, and so forth [Carley et al., 2002]. Table 2.1 provides a simplified representation of the meta-matrix concept.

Contained within each of these single-, multi-relational, and multi-dimensional network constructs are the two fundamental items of interest: the relational ties and actors. Brass offers general summary of social network measures involving the relational ties, the actors, and the overall consequences of the network topology as a result of both [Brass, 1995]. These summaries are shown in Tables 2.2, 2.3,

Table 2.2: Link Attributes [Brass, 1995, pg. 45]

Attribute	Definition
Indirect links	Path between two actors is mediated by one or more others
Frequency	How many times or how often the link occurs
Stability	Existence of link over time
Multiplexity	Extent to which two actors are linked together by more than one relationship (linkages between two given actors occur within several contexts)
Strength	Amount of time, emotional intensity, intimacy, or reciprocal services (frequency or multiplexity often used as measure of strength of tie)
Direction	Extent to which like is from one actor to another
Symmetry	Also referred to as reciprocity, the extent to which relationship is bidirectional

and 2.4. Note that essentially all of the link and actor attributes have associated mathematical definitions and measures. However, only those equations required as part of the methodological development in this research will be presented. For a complete review of SNA-specific formulations, Wasserman and Faust [1994] and Monge and Contractor [2003] serve as comprehensive references.

Despite the fact that a majority of sociological studies and topologically-dependent measures typically approach relationship in a dichotomous matter, it can be seen from Table 2.2 that links may clearly hold more information than a 'yes or no' response within a given context. Additionally, as Renfro [2001] and Sterling [2004] have mentioned, the dichotomous approach to link assessment results in a non-metric measure when the relationships these links capture are asymmetric, or 'one-way.' A metric space is defined as follows.

Definition 6. *A metric space is a nonempty set M of objects (called points) together with a function d from $M \times M$ to $\mathbb{R}$ (called the metric of the space) satisfying the following four properties for all points x, y, z in M:*

1. $d(x, x) = 0$

2. $d(x,y) > 0$ *if* $x \neq y$

3. $d(x,y) = d(y,x)$

4. $d(x,y) \leq d(x,z) + d(z,y)$ [Apostol, 1974, pg. 60-1].

If the sociomatrix and its constituent relationships are symmetric, the values within $\mathbf{X}$ form a discrete metric space [Apostol, 1974, pg. 61]. However, if the relationships are asymmetric, Property 3 and possibly Property 4 of Definition 6 are violated. Nonetheless, as proven by Renfro, non-metric estimates of interpersonal strength still meet the assumptions of classic linear network flow models and their multi-commodity extensions; however, this effect can pose some challenges when dealing with generalized forms of linear flow models [Renfro, 2001, pg. 89-91]. These models and the challenges associated with them are addressed later.

The concept of multiplexity is intriguing for a variety of reasons. Relationships maintained and enforced in multiple contexts offer a potential means for measuring relationship strength, as noted by Brass, as well as its inherent multi-dimensional and subsequent multi-layered approach. Other than the work of Gould [1991], who determined that the solidarity displayed within insurgent ranks was due to the combination of pre-existing informal ties with formal, organizational ties, Monge and Contractor state that "multiple relations on the same set of nodes are quite rare in the research literature" [Monge and Contractor, 2003, pg. 296]. They further explain that ...

> most network researchers believe that many networks are predictive of other networks, that communication networks, for example, are likely to be highly predictive of friendship networks. Even more obvious is the fact that autoregressive networks-the same network at previous points in time-like other autoregressive processes, are likely to predict current values of the network. But until these relations are demonstrated with empirical research using valid statistical procedures, they remain in the realm of speculation [Monge and Contractor, 2003, pg. 296].

Interestingly, Wasserman and Faust recommend against aggregating multiple relations into a single sociomatrix unless "there are strong substantive reasons" for

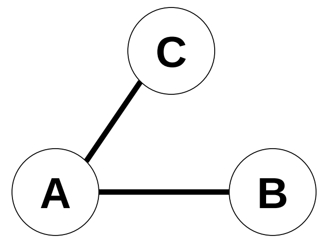


Figure 2.1: Forbidden Triad [Granovetter, 1973, pg. 1363]

doing so [Wasserman and Faust, 1994, pg. 219]. It is an underlying assumption within this research that there are valuable insights to be gained by investigating not only each of the relations, but their combined effect, similar to the work presented by Clark [2005]. However, data collected on non-cooperative networks may not yield graphs that are connected in each layer. Unconnected graphs often prove to be problematic when trying to calculate centrality and reachability indices.

An alternative perspective on the combining of relationships is proposed by Granovetter, who suggested that "the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another" [Granovetter, 1973, pg. 1360]. He further defines the strength of a tie as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" [Granovetter, 1973, pg. 1361]. How such linear combinations are developed remains of theoretical interest. Clark [2005] used a normalized version of the multidimensional centrality measures between relational graphs, presented by Bonacich et al. [2004] (discussed in detail later), as a proxy for contextual weighting. Application of this measure, however, is dependent upon symmetric, unvalued relationship data that comprises a connected graph. In general, this area poses several opportunities for investigating trade-offs between aggregate and independent analyses of contextual relationships, as well as weighting techniques to facilitate aggregation as needed.

Other areas of interest derived from the concept of multiplexity are due to implications of strong and weak ties posed by Granovetter [1973]. For example, consider the relationships depicted in Figure 2.1. Given a strong tie between actors (A) and (B), if actor (A) also has a strong tie with actor (C), then it is unlikely that there exists no tie between actors (B) and (C) [Granovetter, 1973, pg. 1363]. As an example, suppose (B) is a male with a strong relationship with female (A) such that marriage is imminent. Further suppose that (C) represents a member of (A)'s family. Ultimately, it is inevitable that (B) will meet and establish a relationship with (C). The amount of time (A) spends with (B) and (C) separately ultimately leads to a condition of social 'pressure' which can only be alleviated by (B) and (C) establishing a relationship.

On the other hand, the concepts of strong and weak ties studied by Granovetter [1973] had the underlying assumption of normal, interpersonal interaction, as opposed to intentionally surreptitious relationships of people involved in dangerous and anti-social behavior. A classic example could include a male (A) that is married to (B) but is also romantically involved with (C). In order for (A) to proceed (successfully) with this duplicitous behavior, keeping actors (B) and (C) ignorant of each other's existence is likely required. A similar relationship may be desirable in a clandestine network for the purposes of organizational security and therefore must be considered. Such ties are beneficial in limiting exposure to the remainder of the organization if one individual is caught and interrogated. However, as Krebs surmised, maintaining secrecy of organizational ties and operational activities is socially costly [Krebs, 2002].

Complete relational triads or relationships that are comprised of a 'linear combination' of the aspects described earlier, serve as strong ties according to Granovetter [1973]. Alternatively, weak ties are composed of casual or intermittent relationships; however, weak ties are potentially strong themselves. The strength of a weak tie lies in its ability to bridge communication or influence between two or

more distinct groups, or promote diffusion of influence and ideas between them [Granovetter, 1973, pg. 1363-7]. This relationship may prove valuable militarily and is discussed within the next category of measures, actor attributes [Granovetter, 1973, pg. 1364-5].

Table 2.3 summarizes actor attributes commonly used in the SNA literature. Note that within the SNA literature, 'actor attributes' refers not only to attributes specific to that actor (gender, age, and education, for example), but also those attributes that are a direct result of network topology. The latter category characterizes each actor's location and connectivity to adjacent and/or all other actors within the organization.

Actor roles, shown in Table 2.4, are of particular interest in the context of military and national security operations. For example, a bridge may be a member that has a highly-specialized skill and performs critical services, the development of biological weaponry, perhaps, for a number of the network's cliques or cells. Liaisons may comprise the senior leadership and coordination, whereas gatekeepers could be mid-level leadership. Isolates with no links will likely comprise new suspects undergoing investigation; isolates with few links may be indicative of the next suicide bomber, the leader practicing good operations security (OPSEC), or the 'agent in place' awaiting activation to execute instructions established long before surveillance began. As Krebs noted in his analysis of the 9-11 terrorist group, "Those who were trained to fly didn't know the others. One group of people did not know the other group" [Krebs, 2002, np]. In order for the network to function in such a coordinated fashion, a few select individuals had to occupy such roles. This suggests that isolates with few ties may be just as much of interest to intelligence, military, and law enforcement agencies as the most central actors. Unfortunately, this paradigm results in extremely large networks, since an appropriate 'cut-off' point is oftentimes unavailable. Nonetheless, identification of the individuals performing such 'structural'

Table 2.3: Actor Attributes [Brass, 1995, pg. 46]

Attribute	Definition
Degree	Number of direct links with other actors
In-degree	Number of directional links to the actor from other actors
Out-degree	Number of directional links from the actor to other actors
Range (Diversity)	Number of links to different others (others are defined as different to the extent that they are not themselves linked to each other, or represent different groups or statuses)
Closeness	Extent to which an actor is close to, or can easily reach all the other actors in the network. Usually measured by averaging the path distances (direct and indirect links) to all others. A direct link is counted as 1 whereas indirect links receive proportionately less weight.
Betweenness	Extent to which an actor mediates, or falls between any other two actors on the shortest path between those two actors. Usually averaged across all possible pairs in the network.
Centrality	Extent to which an actor is central to a network. Various measures (including degree, closeness, and betweenness) have been used as indicators of centrality. Some measures of centrality weight an actor's links to others by the centrality of those others
Prestige	Based on asymmetric relationships, prestigious actors are the object rather than the source of relations. Measures similar to centrality are calculated by accounting for the direction of the relationship (i.e., in-degree)

Table 2.4: Actor Roles [Brass, 1995, pg. 46]

Role	Description
Star	An actor who is highly central to the network
Liaison	An actor who has links to two or more groups that would otherwise not be linked, but is not a member of either group
Bridge	An actor who is a member of two or more groups
Gatekeeper	An actor who mediates or controls the flow (is the single link) between one part of the network and another
Isolate	An actor who has no links, or relatively few links to others

roles may serve as either an initial target set for influence or application of more elaborate intelligence measures.

Other roles include *representative*, *itinerant broker*, and *coordinator* are described by [Degenne and Forsé, 1999, pg. 128-30]. These . The *representative* is similar to a *liaison*; however, the *representative* takes a more active role in portraying the network to which they claim membership. The *itinerant broker* "facilitates intra-group communication," acting as a third-party that assists two or more groups in achieving common goals. An example would include an outside consultant that mediates or streamlines the operations of several divisions within a company. The *coordinator* is similar to the *itinerant broker* but is part of the two or more networks involved in communication and therefore not acting as a third party [Degenne and Forsé, 1999, pg. 129-30]. The articulator roles from both Degenne and Forsé [1999, pg. 129] and Brass [1995] are depicted in Figure 2.2. The *Ego*, the person or group of interest or focus within the given social context, in the diagrams indicates the structural relationship between two or more individuals, subsets of a given network or entirely separate networks given a specified position.

The final set of measures (see Table 2.5) summarized by Brass assesses the global structure of the network. These are also descriptive measures, however, some concepts may prove useful in developing courses of action intended to affect network performance on a global scale. Examples of interesting uses of these global network

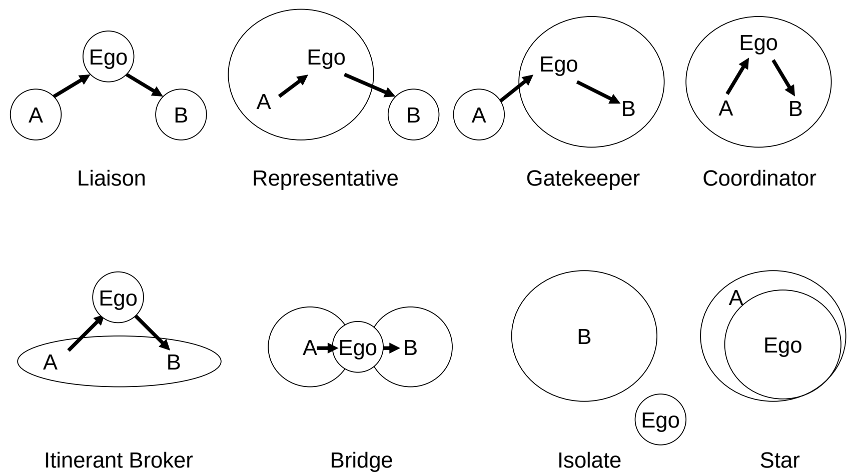


Figure 2.2: Articulator Roles

characteristics may include actions taken to increase the number of components within the network. For example, if the negation of a few select links or actors results in the separation of various terrorist cells, planning and coordination of future attacks may be more difficult for them. The concept and expectation of transitivity, considering Granovetter's forbidden triad, may offer insight into where additional links should exist based upon current data, but not explicitly revealed due to the surreptitious nature of the target network. This essentially leads to the notion that target sets must be developed.

Recent research effort by Borgatti, in an attempt to identify key players within a network, provides one means to identify potential entry points within a non-cooperative network. Borgatti defines two 'key player problems' (KPP).

Definition 7. (KPP-1) *Given a social network, find a set of k nodes (called a kp-set of order k) which, if removed, would maximally disrupt communication among the remaining nodes.*

Definition 8. (KPP-2) *Given a social network, find a kp-set of order k that is maximally connected to all other nodes* [Borgatti, 2003a, pg. 241].

Table 2.5: Network Attributes [Brass, 1995, pg. 47]

Measure	Definition
Size	Number of actors in the network
Inclusiveness	Total number of actors in a network minus the number of isolated actors (not connected to any other actors). Also measured as the ratio of connected actors to the total number of actors.
Component	Connected subset of network nodes and links. All nodes in the component are connected (either direct or indirect links) and no nodes have links to nodes outside the component.
Connectivity	Also referred to as reachability, the extent to which actors in the network are linked to one another by direct or indirect ties. Sometimes measured by the maximum, or average, path distance between any two actors in the network.
Connectedness	Ratio of pairs of nodes that are mutually reachable to total number of pairs of nodes
Density	Ratio of the number of actual links to the number of possible links in the network.
Centralization	Difference between the centrality scores of the most central actor and those of other actors in a network is calculated, and used to form ratio of the actual sum of the differences to the maximum sum of the differences
Symmetry	Ratio of number of symmetric to asymmetric links (or to total number of links) in a network
Transitivity	Three actors (A, B, C) are transitive if whenever A is linked to B and B is linked to C, then C is linked to A. Transitivity is the number of transitive triples divided by the number of potential transitive triples (number of paths of length 2).

Borgatti also relates these problems to those of military interest. For example, KPP-1 would allow target selection in the classical sense. "Given a network of terrorists who must coordinate in order to mount effective attacks, and given that only a small number can be intervened (e.g., by arresting or discrediting), which ones should be chosen in order to maximally disrupt the network?" [Borgatti, 2003a, pg. 241] Note that network disruption is slightly different than previous efforts targeting physical networks. Specifically, KPP-1 seeks not only to break the network into as many components as possible, but seeks resulting components that are as fragmented as possible, all via the selection and removal of the fewest nodes. Related network disruption research has primarily relied upon cut-sets that disconnect a source and sink node [cf. Leinart, 1998; Leinart et al., 2002; Curet et al., 2002]. For KPP-2, the underlying premise is to find a set of actors that would facilitate "the diffusion of practices or attitudes ..." which, militarily, "translates to locating an efficient set of enemies to surveil, turn (into double-agents), or feed misinformation to" [Borgatti, 2003a, pg. 241].

The rationale for the development of the KPP methods lies in the fact that many centrality measures were not developed with the intention of (potentially) adversely affecting the network under study. Instead, social scientists attempted to observe, and occasionally predict, behavior. Consequently, the measures that evolved do not necessarily translate to network disruption or 'seeding,' especially if the objective is to select the optimal set of individuals ($k > 1$) that accomplishes these objectives [Borgatti, 2003a, pg. 241-7].

To address these issues, Borgatti develops a heuristic to analyze both problems, but notes that the solutions are inherently "considerably less than optimal" [Borgatti, 2003a, pg. 247]. Additionally, it does not appear that the measures of 'goodness' (of the solutions) used within the heuristic can account for directed arcs. Variations and applications of classical covering, partitioning, and p-median problems, discussed later, address some of these issues [cf. Nemhauser and Wolsey, 1999]. The end

product is that, as opposed to the heuristic approach currently in use, the application of mathematical programming techniques are guaranteed to provide the optimal solution or solutions. However, knowing which actors to target, based upon the KPP results, implicitly assumes that those actors are accessible.

In their analysis of other network disruption algorithms, Degenne and Forsé [1999] discuss the actors' vulnerability as a function of "the risk of becoming isolated if one or more individuals drop out of the relation under study" [Degenne and Forsé, 1999, pg. 103]. This implies that the higher value targets–those that are more likely to disrupt (disseminate misinformation) if they are removed (influenced), respectively–are those with high betweenness centrality and are also those that are potentially more difficult to reach. This concept is illustrated in Figure 2.3.

Note that 'high-value targets' in this context does not necessarily imply a key leader within an organization (e.g., Osama bin Laden is considered a high-value target within Al Qaeda). The concept of 'high-value' in this case refers to the potential damage, via destruction or dissemination of misinformation, incurred by removing or co-opting a particular individual or set of individuals. An example could include the individual that assembles bombs for the local cell's operations. For KPP-1 in particular, the trade-offs between resources required to successfully engage a highly central, but highly valued, target versus a more easily engaged, but less valued, target must be assessed. This approach, however, implicitly assumes centralized control of the network and its activities, which is unlikely the case for semi-autonomous cells operating within today's terrorist networks.

The KPP-2 concept is important in that the injection and subsequent spread of influence (e.g., PSYOP or influence operations) requires an entry point into a network with "formidable barriers to entry and exit" [cf. Post, 2005; Rothenberg, 2002, pg. 39]. Rothenberg notes that "The breakdown of such a network, whether on the local or global scale, depends obviously on two factors: money and trust. The US and other governments are in hot pursuit of the former, but appear . . . befuddled by

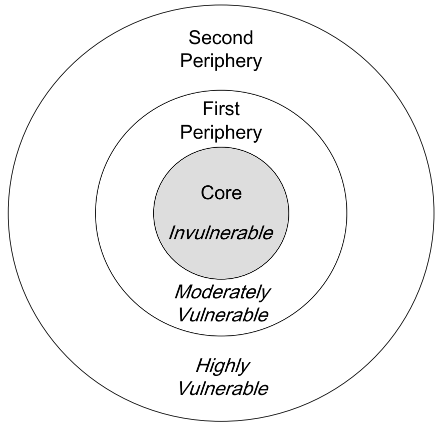


Figure 2.3: Core-periphery partitioning [Degenne and Forsé, 1999, pg. 103]

the latter" [Rothenberg, 2002, pg. 40]. With relatively low-technology tactics such as suicide bombers and improvised explosive devices becoming terrorists' weapons of preference, money may simply be an enabler or perhaps not even necessary for some operations. Trust, and the corresponding relationships it nurtures, appears to be the main strength of today's terror networks. For example, the second lesson of the Al-Qaeda Training Manual is entitled *Necessary Qualifications and Character for Organization's Members* and details the requisite commitment, attributes, and willingness to undergo martyrdom [Post, 2005, pg. 25-32]. Consequently, techniques and theory that contribute to attacking trust are appealing. With strength due to trust in mind, Rothenberg puts forward a view, in opposition to this research, summarizing the challenges associated with such an effort.

> Sowing viruses of distrust is difficult within a network that has major obstacles to entry, is highly decentralized, and whose leader's status as a symbol is likely to be untouched by what will happen to him. Peripheral persons play a role, but primarily as purveyors of needed goods and not as participants; such roles may or may not be useful in infecting the network [Rothenberg, 2002, pg. 40].

In the absence of actual terrorist networks to conduct experiments in this area, the influence-related aspects of the methodology proposed in Chapter III may not be readily verifiable in a formal sense. However, such points do highlight the need to ac-

count for operational risks when dealing with non-cooperative organizations. Lastly, Rothenberg notes the complexity of data collection for typical network studies as requiring numerous interviews, ascertaining the levels or strengths of relationships, and assessing network dynamics [Rothenberg, 2002, pg. 36]. These efforts are even more problematic when dealing with non-cooperative, adversarial, and adaptive networks that cannot be easily revealed through open and direct inquiry.

2.2.3 The Challenges of Network Data

Methods traditionally used to collect sociometric data include questionnaires, interviews, observations, archival records, experiments, and others; this implies that data sets comprise populations rather than subsets of them [Wasserman and Faust, 1994, pg. 45]. Granovetter notes that "It is clear why network methods have been confined to small groups: existing methods are extremely sensitive, in their practicality, to group size because they are population rather than sampling methods" [Granovetter, 1976, pg. 1287-8]. Due to the potential $n(n-1)$ number of directed ties between n individuals, collecting complete and accurate data on large populations is costly and problematic. Further, Granovetter argues that such studies can only make implicit connections between the nature of the data collected and the nature of the true population from which the data came. Ultimately, "...we are left guessing about the representativeness of the patterns of social relations found" [Granovetter, 1976, pg. 1288].

Unless the individuals that comprise the population are known with certainty, how representative the sample will be of the true population will always remain in question. For example, Tsvetovat and Carley [2005] have estimated Al Qaeda membership to be as high as 120,000. Even if such an extensive network could be mapped, it is likely that the magnitude would leave most current analysis capabilities computationally intractable. Hence, samples or subsets of the true networks comprise currently available data sets.

Other issues pervading network data that describes non-cooperative networks includes missing data and potential structural bias as a result of the data gathering processes available. In order to truly capture the information regarding a relationship between two individuals, both individuals must be questioned [Stork and Richards, 1992, pg. 194]. When dealing with terrorists, unless both individuals are in custody and amenable to truthful interviews, this is a difficult process dependent upon the skills of both the interviewer and the interviewee, as well as some degree of luck. As a result, analysis must be performed on incomplete data.

Robustness of classical network centrality measures given data errors such as "...edge deletion, node deletion, edge addition, and node addition" has been explored by [Borgatti et al., 2006]. Unfortunately, the underlying graphs used in their experiment were random in nature, as opposed to a more representative small-world network topology. Previous research suggests that random networks, even when degree distribution is carefully controlled, are not always representative of social networks due to "...non-random social phenomena at work in the shaping of the network" [Newman et al., 2002, pg. 2571]. Nonetheless, the authors concluded that responses to error were ultimately a function of error type and network density [Borgatti et al., 2006]. Although similar findings using 'real' and experimental network data are provided in Bolland [1988], the redundant nature of the data may have biased the experimental results. The sensitivity of other, more general, network measures such as global efficiency, critical path length, density, diameter, and radius of scale-free graphs has also been explored by [Thomason et al., 2004]. Whether these conclusions map to more appropriate network topologies remains to be seen and, based upon the analysis of network disruption as seen in Albert et al. [2000], is likely heavily dependent upon where the missing data lies within the network.

As aforementioned, social network studies typically deal with populations as opposed to samples of a population, for example, all the children in a classroom [Moreno, 1953], the tribe members occupying a chain of islands [Bonacich et al.,

2004], family members, and company managers [Wasserman and Faust, 1994, pg. 738] to highlight a few. Three methodologies appear to dominate the literature regarding the capture of network data. The first two methods, *snowball sampling*, also referred to as expanding selection, and *fixed list*, or *fixed selection*, assume that respondents are either somewhat willing or can be persuaded to provide relational information. The third, *targeted sampling*, was developed specifically for the study of transmission of AIDS among intravenous drug users. The fact that the members of such a network were involved in illicit drug use essentially results in a clandestine network not totally unlike that of a terrorist organization [Watters and Biernacki, 1989].

Snowball sampling procedures are defined by a predetermined number of s stages and k names. The steps of the methodology are shown in Table 2.6. Note that if the 'random sample' in step 1 is replaced with 'detected or detained set,' this procedure is essentially the methodology used by law enforcement and intelligence agencies, with s continuing either indefinitely or until the entire group is discovered and no longer a threat. However, the respondents in this case, if captured, are generally unwilling to provide information. Consequently, the approach is similar but the results may remain limited in comparison to willing responses or biased due to the reliance upon deceptive information. If information is gathered surreptitiously (for example, via wiretaps or other forms of electronic monitoring), then unless the target individuals are cognizant of the surveillance or practicing some form of operations security, they could be considered 'willing' in the sense that they are not intentionally withholding information. The resultant search pattern, however, will likely be limited and more sparse compared to an open study simply due to the inherent security needs of the clandestine organization.

Fixed list sampling entails the provision of a list of other people and getting the respondents to indicate which ones on the list they consider themselves sharing some level of relationship within the context under investigation. This is accomplished for

Table 2.6: Snowball Sampling Procedure [Goodman, 1961, pg. 148]

Step	Description
1.	A random sample of individuals is drawn from a given finite population.
2.	Each individual in the sample is asked to name k different individuals.
3.	The individuals who were not in the random sample but were named by individuals in it form the first stage.
4.	Each of the individuals in the first stage is then asked to name k different individuals.
5.	The individuals who were neither in the random sample nor in the first stage but were named by individuals who were in the first stage form the second stage.
6.	The procedure is continued until each of the individuals in the sth stage has been asked to name k different individuals.

at least all respondents that appear on the original list [Doreian and Woodward, 1992, pg. 217-18]. In an interrogation setting, detainees may be given such a list and asked to confirm relationships or are perhaps questioned over time until variations and inconsistencies arise. These indicators can then be used to draw further information from the subject.

The most noticeable, implicit assumptions for these methods includes: the researcher must know at least a few of the members of the network; the respondents are accessible for interview; and, the respondents answer in a truthful manner. All of these assumptions prove problematic when dealing with non-cooperative networks. In addition, structural biases may be introduced as a result of sampling technique. For example, Doreian and Woodward [1992] found that between snowball and fixed list techniques applied to the same target network, "(1) the [actors] included in the two selection procedures differ; (2) network-based [measures] differ; (3) the substantive contents of the included ties differ; and (4) the structure of the networks differ" [Doreian and Woodward, 1992, pg. 216]. Although both snowball and fixed list sampling techniques could be viewed as similar in nature to the interrogation process, they must begin with a 'captive' respondent. As noted in Watters and Biernacki

however, the responses of captured members of non-cooperative networks and the subsequent network and member characteristics developed may not be representative of the network that exists outside of the controlled and persuasive environment of incarceration [Watters and Biernacki, 1989, pg. 417].

Perhaps the most concrete evidence of a sociometric sampling technique successfully resulting in an 'accurate' representation of the true (population) network is the *targeted sampling* procedure posed by Granovetter [1976]. He suggests "given a population of size N, the method proposed is to take a number of random samples from that population, each of size n (with replacement), and within each such sample ask each respondent some sociometric question about each other respondent" [Granovetter, 1976, pg. 1290]. Granovetter proves that this approach yields an unbiased estimate of the true network's density [Granovetter, 1976]. Underlying assumptions are, once again, that network members are known, respondents are willing to participate, and relations are symmetric (despite the fact that he recommends that respondents be questioned both ways). Granovetter does account for one-way questioning (that is, relying more heavily on the assumption of symmetry) but does not discuss to the statistical implications of this approach upon his estimate [Granovetter, 1976, pg. 1297]. Further, guidelines for the parameters accounting for "the number of samples taken and the size in each sample" are provided in a similar fashion to the determination of sample size to meet certain uncertainty criteria [Granovetter, 1976, pg. 1290-95]. However, this again assumes that the true population membership is known and that the necessary number of samples of the same size of n individuals is feasible.

Although there are similarities between cooperative and non-cooperative data gathering, the latter organization will always 'fight back' against investigators, providing either no information or possibly misinformation, the latter of which may in some cases still be useful. These issues will continue to plague decision-makers and analysts as long as the target networks are non-cooperative. As opposed to

Table 2.7: Covertness Factors [Tsvetovat and Carley, 2005, np]

1.	Strong religious (in case of Islamic groups) or ideological (in case of Sendero Luminoso and other South American guerilla groups) views that allow members to form extremely strong bonds within a cell.
2.	Physical proximity among cell members, often to the extent of sharing living quarters, working and training together.
3.	Lack of rosters on who is in which cell.
4.	Cell members being given little knowledge of the organizational structure and the size of the organization.
5.	Little inter-cell message traffic.
6.	Information about tasks issued on a need-to-know basis, so very few people within the organization know about the operational plans in their entirety.
7.	Cells are often formed on the basis of familial or tribal ties, or strong interpersonal ties forged in training.

ascertaining the effect upon centrality measures due to errant data, future work may benefit from approaches that maximize the likelihood of discovering a certain, hopefully high, percentage of the true network structure. Unfortunately, simply obtaining data on these adversarial networks is difficult at best. Tsvetovat and Carley describe the primary factors that enable a terrorist network to remain covert; these are provided in Table 2.7. Note that, from a Department of Defense perspective, the concept of covertness presented by Tsvetovat and Carley actually aligns more closely with the concept of 'clandestineness,' which leverages secrecy to mitigating potential damage or interruption of operations due to the exposure of members or activities.

To this point, network topologies, nuances of social networks, and the challenges of collected data characterizing non-cooperative networks have been discussed. However, externally affecting such networks involves much more than the topologically imposed constraints, or opportunities, within which the network members must work. It is also important to understand how these individuals may react to outside sources of influence. The next section discusses open source elements of the psycho-

logical aspects of terrorist decision making, as well as current efforts attempting to model this phenomena.

2.3 The Psychology of Terrorists

The President of the United States defines terrorism as "premeditated, politically motivated violence perpetrated against innocents" [The President, 2006, pg. 5]. To better deal with terrorists and terrorist organizations, an understanding of their underlying psyche, motives, and overall goals is required. This information would, in theory, allow analysts to model the behavior of such individuals and their organizations, thus providing an opportunity to learn, predict, and directly or indirectly thwart this type of threat. Several ongoing efforts attempting to encode behavior in simulated agents representing terrorists are also reviewed. These simulations are undertaken in order to gain insight and to provide courses of action that are more likely to marginalize the threat posed by terrorists.

This section discusses several existing methods that attempt to model the behavior of terrorist groups or, more specifically, the terrorists themselves. Potential benefits of this research thread could include (1) gaining insight into the underlying causes for motivating an individual to engage in these activities; (2) describing the incorporation of these concepts into agent-based simulations for study; and (3) illustrating the use of these simulations to evaluate various courses of action, ranging from close-combat operations to application of other instruments of power and/or international diplomacy. Although the second and third benefits are beyond the scope of this research, the first must be addressed when approaching this problem from either a static or dynamic viewpoint.

Maslow's hierarchy provides a solid basis for describing behavior "...across cultures, age groups, and generations..." [Johns and Silverman, 2001, pg. 4]. However, when considering that physiological and safety needs come before all other

needs, the phenomena of “suicide bomber” presents a dichotomy. Johns and Silverman attempt to partially resolve these types of issues by incorporating emotion into a decision theoretic representation of ‘agent’ behavior. As this work clearly presents a learning opportunity, the decision theory that underlies this effort requires review.

An effort parallel to that of Johns and Silverman [2001] by Silverman et al. [2001] extends the agent-based simulation capability by offering a means to ‘dial-up’ an adversary, inferring a capability to choose an adversary and possibly the situational context with the resulting agent behavior that is tailor-made to accommodate social, ethnic, and other characteristics that may perturb a basic model of behavior. These efforts, of course, beg the question of what aspects and underlying psychological models of decision making, emotion, motivation, and rationality are most suitable to characterize the adversary of interest, the terrorist.

2.3.1 What should be modeled and why?

Given its inherent complexity, modeling human behavior is in itself a daunting task. Modeling human behavior that falls within the realm of terrorist activity may be even more difficult. Some terrorist activities simply fall within what many perceive to be irrational behavior (e.g., suicide bombing) and, very generally speaking, involves an enemy that ascribes to different ideologies than our own. The “irrationality” of suicide terrorism is a prevalent misconception [Driscoll, 2005]. Terrorists are not always motivated by fanatical interpretations of religion and do not always originate from the stereotypical ‘underprivileged’ classes of society. In fact, terrorists are “deeply committed, maintain excellent intellectual ability, planning, problem solving, interpersonal skills and self-discipline” [Driscoll, 2005]. It is true that the U.S. armed forces are willing to sacrifice their lives for our own ideologies (e.g., liberty and democracy). However, unlike the suicide bomber, Western forces tend to focus on the capitulation or death of the adversary with minimal loss of friendly forces rather than intentionally sealing our own troop’s demise during the course of warfare.

Assuming that this phenomenon can be explained, as attempted by the work of Sprinzak [2000], what other aspects of terrorist behavior are of interest? Weaver et al. [2001] and Johns and Silverman [2001] develop models to facilitate military training, particularly in the area of guerilla warfare within a hostile urban environment. The ultimate modeling goal may be to predict how a given threat will behave and consequently understand how to defeat it. From a long-term perspective, improving understanding of terrorist behavior may improve the capability to avoid the underlying conditions that contribute to the recruitment of future generations of terrorists.

2.3.2 Rational Decision Making

There are a variety of definitions of rationality in the literature. For example, in the context of microeconomics, a "rational consumer will choose a market basket where the marginal utility of the last dollar spent on all commodities purchased is the same"–essentially where a budget line is tangent to the highest indifference curve [Mansfield, 1994, pg. 83-4]. In a related, but sometimes more practicable genre, decision theoretic implementation of multi-objective utility analysis assumes that once an "assignment of utility numbers to consequences" is complete, the optimal, and rational, strategy requires selection of the alternative that maximizes expected utility [Keeney and Raiffa, 1993, pg. 7]. Noting that selection of alternatives may also be subject to budget constraints, the decision theoretic approach provides a deeper focus on the derivation of utility. Hence, the economic approach may prove useful in determining levels of humanitarian aid or standard of living improvements required to persuade people from succumbing to the appeal of terrorist organizations, but its role in trying to understand and predict current terrorist behavior remains unknown. The decision theoretic approach certainly lends itself to implementation within a decision model of a simulated actor or agent in agent-based simulations, shown later in this section. However, in the context of both the actions and ideologies

of terrorists and their organizations, optimal decision strategies may vary drastically among individuals. Hence, a more fundamental understanding of human behavior in the context of decision making must be considered. Commonly accepted, albeit Western-focused, models of human behavior, along with its role in decision making, are reviewed.

2.3.3 Maslow's Hierarchy

A. H. Maslow developed the concept of the hierarchy of needs in order to better understand human behavior. His construct has been used extensively in business organizations to assess what may or may not motivate employees to continuously improving levels of performance. The needs are hierarchical in nature and are summarized in Table 2.8. Another important aspect of this theory is its underlying assumptions. Costley et al. summarizes these as follows:

1. Motives are highly complex, and no single motive affects behavior in isolation. A number of motives are always in operation at the same time.
2. There exists in each individual a hierarchy of needs that requires, in general, that lower-level needs must be partially satisfied before higher-level needs affect behavior.
3. A satisfied need is not a motivator. When a need is satisfied, another need emerges, so that the individual always remains in a motivated state. Higher-level needs can be satisfied in a greater variety of ways than lower-level needs [Costley et al., 1994, pg. 219].

Considering that *Physiological* and *Safety* are two of the most basic needs (and therefore, generally take priority over all others), it may be hypothesized that once these are met, the stage is set to fulfill needs generated within the higher levels. However, in the context of suicide bombing, the obvious question becomes "Would an individual guarantee the ruin of the most basic needs (*Physiological* and *Safety*)

Table 2.8: Needs Hierarchy [Costley et al., 1994, pg. 219]

Urgency	Need	Examples
Most Urgent	Physiological	Food, water, rest, and shelter
	Safety	Security and protection
↓	Social	Belonging, acceptance, and friendship
	Esteem	Recognition, status, and self-esteem
Least Urgent	Self-Actualization	Creativity and self-realization

in order to achieve a need in the higher level of the hierarchy?" Another theory, developed by Alderfer [1969], offers a potential explanation of this dichotomy.

2.3.4 Existence, Relatedness, Growth Theory

Existence, Relatedness, Growth (ERG) Theory was developed to address some of the questions surrounding Maslow's hierarchy. ERG differs from Maslow's theory in two major areas, specifically its structure and the assumptions linking the structure. As opposed to the needs categories defined by Maslow, Alderfer clusters needs into the three categories: Existence, Relatedness, and Growth. *Existence* needs account for material and physiological desires, *Relatedness* includes relationship needs of significant others, and *Growth* accounts for the "creative or productive effects on himself and the environment" [Alderfer, 1969, pg. 145-6]. Clearly, the structure mirrors Maslow's to some extent.

However, this particular classification scheme avoids some of the overlap problems suffered by Maslow's hierarchy [Alderfer, 1969, pg. 147]. Additionally, Alderfer's hierarchy is not strictly ordered; consequently, "it does not assume lower-level satisfaction as a prerequisite for the emergence of higher-order needs" [Alderfer, 1969, pg. 142]. It is this underlying premise that offers a means to explain a terrorist's Jihadist aspirations that essentially guarantee their own demise. Lastly, Alderfer notes that these categories of needs "...provide the basic elements in motivation" [Alderfer, 1969, pg. 145].

2.3.5 Motivation-Hygiene Theory

A third, relatively mainstream, theory of motivation is due to Herzberg et al. [1965]. In an attempt to find the key factors that lead to job satisfaction, Herzberg et al. developed the theoretical bases for a person's attitude toward their job [Herzberg et al., 1965, pg. 3]. Factors associated "...with conditions that surround the doing of the job," thereby affecting the psychological health of the work environment, comprise the "factors of hygiene" [Herzberg et al., 1965, pg. 113]. Fulfillment of such factors may prevent dissatisfaction but do not necessarily guarantee job satisfaction. Hence, the second basis of the theory deals with motivation factors.

Motivation factors are those that "...lead to positive job attitudes [because] they satisfy the individual's need for self-actualization in his work" [Herzberg et al., 1965, pg. 114]. Both factors ultimately serve the psychological needs of the worker, "but it is primarily the motivators that serve to bring about the kind of job satisfaction and ...improvement in [future] performance" [Herzberg et al., 1965, pg. 114].

Costley et al. compare and contrast Herzberg et al.'s theory with that of Maslow's. Essentially, the motivation factors coincide with self-realization and esteem needs, whereas the maintenance or hygiene factors coincide with social, safety, and physiological needs [Costley et al., 1994, pg. 247]. This approach to modeling the psychological underpinnings of motivation and rationale for behavior was developed for, and is continuously applied to, business-employee settings. Although an abstraction of this theory could be applied to the underlying factors in terrorist behavior, such an attempt in this specific area is currently beyond the scope of this research. However, considering each of these theories attempts to explain what motivates individuals while also accounting for emotional states in one form or another, a complete explanation of suicidal terrorist behavior is likely dependent upon a combination of both motivation and emotion.

2.3.6 *Motivation*

Decision theory states that people base their decisions upon assumptions and information that describe which of their choices results in the most favorable outcome. Costley et al. notes that several expectancy theories of motivation all share one underlying assumption "…that people choose behaviors based on their expectations about the outcomes" [Costley et al., 1994, pg. 232]. This expectation and consequent measure of motivation is a "function of the value placed on potential rewards (referred to as valence) and the perceived probability that the effort will be successful (referred to as expectancy)" [Costley et al., 1994, pg. 232]. Given that different ideological differences may result in different concepts of what is a 'reward,' this implies that in order to ascertain a terrorist's behavior, it must be done so through their perspective and not our own.

2.3.7 *Emotion*

Arguments for incorporating emotion within decision making are discussed in Ellis and Hunt [1993] and Rolls [2001]. Ellis and Hunt state that "…it is apparent that emotional or affective states can very much influence cognitive processes in important ways" [1993, pg. 333]. Unfortunately, they then illustrate several studies that have shown that emotional states (e.g., depression) do not necessarily impede an individual's ability to make contingency judgments [Ellis and Hunt, 1993, pg. 353-55]. Counter to this argument, Rolls suggests that "Emotions can usefully be defined as states produced by rewards and punishers. Rewards are stimuli for which (a human) will work, and punishers are stimuli that (a human) will work to escape from or avoid" [Rolls, 2001, pg. 4444]. Consequently, these phenomena are "…reinforcers in that they alter the probability of behavior" [Rolls, 2001, pg. 4444]. That is, despite the valence and expectancy values described by Costley et al., the resulting motivation value may be changed or influenced by emotions or the external conditions that initiated the emotional state and level. This concept is illustrated in Figure 2.4

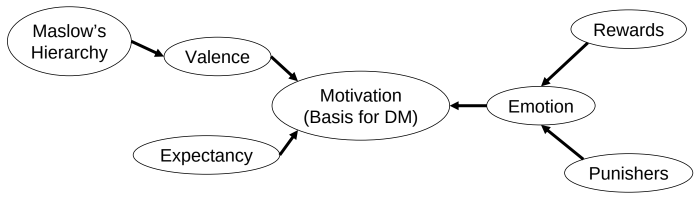


Figure 2.4: Factors Influencing Motivation

and appears to be the underlying premise of the modeling efforts pursued by Johns and Silverman [2001].

Rolls summarizes the functions of emotion. Of these functions, several are of interest in the context of modeling terrorist decision-making. The first function is that emotion serves to elicit certain autonomic and endocrine responses, preparing the body for action [Rolls, 2001, pg. 4444]. A classic example of this function is an athlete getting psyched up using the external stimuli (the crowd's roars, for example) to boost performance. A similar event is the fight or flight response invoked when a human or animal is in immediate and obviously mortal danger.

Rolls also notes that emotion in itself is motivating; previous experiences lead to future actions performed. [Rolls, 2001, pg. 4445] Essentially, levels of fear (or gusto) may influence or facilitate the motivation to achieve a goal. Third, he indicates that "the current mood state can affect the cognitive evaluation of events or memories, and this may have the function of facilitating continuity in the interpretation of the reinforcing value of events in the environment" [Rolls, 2001, pg. 4445]. For example, fear of disappointing Allah, family, comrades, and friends by failing to achieve a terrorist act may allow the individual to focus and continue their mission despite the fact that success will result in their own demise. This is closely related to the next function of emotion–"by enduring for minutes or longer after a reinforcing stimulus has occurred, it may help to produce persistent and continuing motivation and direction of behavior, to help achieve a goal or goals" [Rolls, 2001, pg. 4446].

Finally, "rewards and punishers, and the emotional states they produce, provide a common currency for the behavior selection process between competing alternative actions" [Rolls, 2001, pg. 4446]. This provides further justification for the conceptual model noted in Figure 2.4. Additionally, rationality with regard to the decisions made by terrorists may not fall within the realm of maximizing expected utility, at least from a Western, decision-theoretic perspective. Extremists' definitions of acceptable goals, means, and the underlying emotions and motives employed must be investigated.

2.3.8 What is a rational terrorist?

As previously mentioned, the *modus operandi* of terrorist organizations is to achieve their goals by employing tactics that are "...almost supernatural, extremely lethal, and impossible to stop..., " particularly in the case of suicide bombers [Sprinzak, 2000, pg. 66]. Tactically, this method is advantageous in that ...

> [Suicide bombing] is a simple and low-cost operation (requiring no escape routes or complicated rescue operations); it guarantees mass casualties and extensive damage (since the suicide bomber can choose the exact time, location, and circumstances of the attack); there is no fear that interrogated terrorists will surrender important information (because their deaths are certain); and it has an immense impact on the public and media [Sprinzak, 2000, pg. 66-8].

Essentially, Sprinzak suggests that martyrdom has become the terrorists' primary option against opponents commanding extremely capable military forces. The notion of suicide bombers is not new, and dates back to at least "as early as the 11th century; the Assassins, Muslim fighters living in northern Persia, adopted this strategy to advance the cause of Islam" [Sprinzak, 2000, pg. 68]. However, martyrdom has remained an underlying requirement for this strategy, but it is "...not merely the product of religious fervor," and it varies "...not only by culture, but by circumstance" [Sprinzak, 2000, pg. 68]. To further complicate the problem of modeling this phenomenon, Sprinzak's review concludes that there is "no single psychological

or demographic profile of suicide terrorists . . . , but several types of people with the potential willingness to sacrifice themselves for a cause" [Sprinzak, 2000, pg. 68].

Emotions play a key role in leveraging these individuals; specifically, exploiting religious beliefs and the rewards of an afterlife, ". . . patriotism, hatred of the enemy, and a profound sense of victimization" [Sprinzak, 2000, pg. 69]. With this in mind, Sprinzak suggests that the more useful exercise is analyze the leaders, as opposed to the bombers, that choose this strategy, and indicates that "leaders who opt for this type of terrorism are usually moved by an intense sense of crisis, a conviction in the effectiveness of the tactic, endorsement by the religious or ideological establishment, and the enthusiastic support of their community" [Sprinzak, 2000, pg. 69-70]. It is these situations that may lend themselves to modeling. For example, application of various instruments of power or international support may erode the endorsement and support provided to the terrorist organization. With this in mind, the issue of mapping such abstract concepts as emotions, their link to motivation, and the resulting actions, decisions, and behaviors of terrorist organizations and the individuals that comprise them is of interest.

2.3.9 Realistic Model of Rationality

Obviously, the task of modeling human behavior and decision-making, particularly of non-cooperative individuals, is no simple undertaking. Slade notes that the traditional economic model of rationality requires a great deal of data that is likely unavailable; even so, if all the data is available, the approach of calculating the optimal decision may be intractable [Slade, 1995, pg. 126]. In an attempt to develop a more realistic and implementable model of a rational decision-maker, Slade's work furthers the concept of bounded rationality developed by Simon, "which incorporates information processing constraints in an effort to reflect the limitations of human cognition" [Slade, 1995, pg. 126]. As opposed to bounded rationality which presumes a decision-maker must satisfice rather than optimize, Slade's model

Table 2.9: Assumptions [Slade, 1995, pg. 126-7]

-	An agent (i.e., an individual) has many goals with varying preferences; some goals are more important than others.
-	An agent executes plans (a sequence of actions) in order to achieve specific goals; this behavior requires resources.
-	An agent has limited resources (e.g., time, money, and cognitive capabilities).
-	Different agents have different goals and resources; decision-making is subjective.
-	An agent allocates resources to achieve their preferred goals.
-	Since knowledge is considered a resource, an agent is not irrational if they fail to achieve a goal due to lack of knowledge.
-	Emotions are a reflection of goal states.
-	An agent has relationships, positive and negative, with other agents, with varying strengths.
-	Through a relationship, an agent adopts the goals of the other agent with a preference related to the strength of the relationship.
-	Decisions require justification.

requires the decision-maker to justify choice, thus offering a means to incorporate emotions and motivations with extremists' decision making processes [Slade, 1995, pg. 126]. The underlying assumptions for this model are listed in Table 2.9.

These assumptions were used to develop a decision-making model that explicitly accounts for the "representation of goals, choices, relationships, strategies, and the use of natural language to produce explanations ..." which comprise the justifications for a given decision [Slade, 1995, pg. 127]. It is important to note that Slade expanded the traditional focus from a single decision-maker to one that viewed "decision-making as a social process through the adoption of goals from interpersonal relationships" [Slade, 1995, pg. 129]. Slade suggests that the multi-agent phenomena of advice, persuasion, and negotiation can be explored by this framework [Slade, 1995, pg. 129]. Considering that "a suicide terrorist is almost always the last link in a long organizational chain that involves numerous actors," this framework offers a promising modeling approach [Sprinzak, 2000, pg. 69]. In addition, if the adoption of goals is indeed related to the strength of interpersonal relationships,

this fact could be exploited when deciding how to develop and execute psychological operations that seek to influence an organization's goals. However, just as there are data issues associated with the traditional economic and decision theoretic approach (that is, expected utility and single decision-maker assumption, respectively), the non-cooperative nature of these individuals may pose similar problems. Additionally, the psychosocial aspects considered within an appropriate model are complex. Harris provides supporting rationale for modeling the decision-making and behavior (of terrorists) ...

> "We all want to make sense of our world, and at no time more urgently than when our world is suddenly behaving strangely. But in order to make sense of such strangeness, we must be able to reduce it to something that is not strange–something that is already known to us, something we know our way around" [Harris, 2002, pg. 19].

He also noted, however, that the process of understanding the strangeness inevitably leads to an analysis from our own viewpoint, and not necessarily from that of the "culturally exotic" enemy [Harris, 2002, pg. 19]. This likely remains as one of the most challenging issues to future modeling efforts. The next section presents an overview and associated challenges of some current, but not necessarily all-inclusive, agent-based approaches to this problem.

2.3.10 Current Behavioral Modeling Efforts

Today's demand for realistic video games has promoted the development of computer science techniques to model virtual, artificially intelligent opponents. These technologies may serve as a launching point for modeling terrorist behavior. Of particular interest in the context of modeling decision-making is the agent-based approach to simulation. This simulation framework "consists of individual agents, commonly implemented in software as objects...that have states and rules of behavior" [Axtell, 2000, pg. 2]. Axtell also suggests that the efforts to reduce the computational challenges associated with agent-based simulation by either limiting

intra-agent activity or limiting the options available to agents are similar to the concept of bounded rationality. Unfortunately, the implication is that equilibrium to a specific answer via this simulation approach may never be achieved. However, insight is still possible, but one must also consider the accuracy of the input data when relying upon a definitive answer from this type of model [Axtell, 2000, pg. 9].

Silverman et al. [2001] are actively involved in model development, focusing on the behavior and decision-making of terrorists and terrorist organizations. One such effort involves a compilation of Human Behavior Models (HBM). HBM is based upon diverse, authoritative literature, and attempts to "...quantify the impact of human performance to internal and external stressors, and help to capture the role of personality and individual differences (within simulations)" [Silverman et al., 2001, pg. 1]. This approach also incorporates a Performance Moderator Function (PMF), which relates internal individual or group characteristics and their interaction within a given environment to estimate a level of performance or behavior as a result of environmental stressors [Silverman et al., 2001, pg. 1-2]. One of their goals was to facilitate the difficult task of gathering data useful to mathematical or computational implementation. Additionally, since the data was to provide a means to simulate behavior, it had to be collected within the context of a framework. This framework is illustrated in Figure 2.5.

The steps Acquire, Best-fit Situation, Course of Action, and Direct are analogous to the well known Observe, Orient, Decide, and Act (OODA) loop and are described by Silverman et al. as:

- The agent "acquires data and cues from the external world (x, x'); based upon their prior focus of attention (f) and level of arousal or valence (v), and filters out noise to produce a set of state variables (s) as output."
- The agent evaluates the situation based on the cues attended to (s) and patterns recognized from experience, doctrine and value sets, denoted $P(s|H)$.

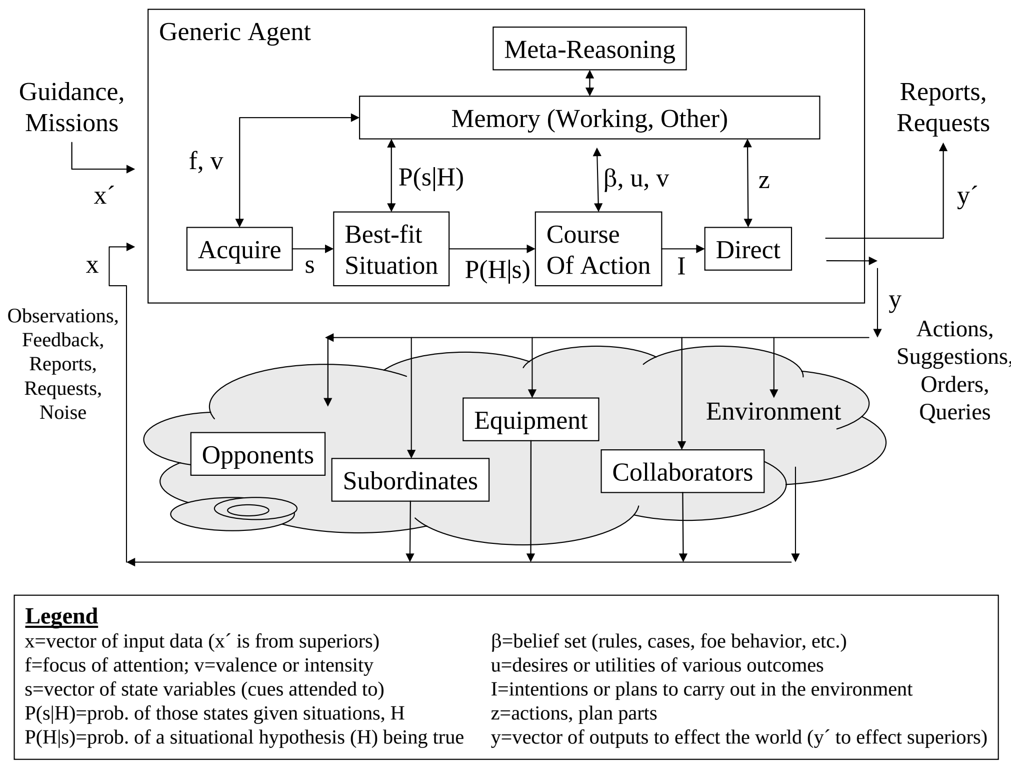


Figure 2.5: Decision-Making Agent [Silverman et al., 2001, pg. 4]

This leads to several plausible hypotheses (H) each with a corresponding likelihood of $P(H|s)$.

- The agent then chooses what course of action to pursue by selecting a decision rule based on doctrine and on the time available to plan and decide. Depending upon the overall objectives and temporal constraints, the agent applies "...utility or desire levels (u), emotional intensity levels (v), and belief sets (β), about the effectiveness of their actions over time and space against the opponent" in order to ascertain the best course of action, denoted the intended plan or intent (I).

- The agent then maps actions required to achieve the plan to output signals, denoted (y, y'), which are enacted by the subsequent steps (z) "...needed to carry out the plan and achieve the intention." The overall effect is to "...optimally control its portion of the external world including its own behavior as well as that of others it might be able to influence" [Silverman et al., 2001, pg. 4].

The other aspect of this framework of interest to this research effort is the *meta-reasoning* block, shown in Figure 2.6. As illustrated, the meta-reasoning process begins with an evaluation of the current emotion level of the agent. Given this influence and information from external and internal sources, the agent tries to decide what course of action best suits the agent's needs, evaluated by the emotion appraiser [Silverman et al., 2001, pg. 5]. It is interesting to note that the authors "reject (Maslow's) concept of seriality of needs" (for example, survival needs must be fulfilled before intellectual fulfillment) [Silverman et al., 2001, pg. 5]. Therefore, this approach favors the theory proposed by Alderfer [1969] and consequently allows for the situation where an individual would place martyrdom or a 'greater cause' before their own mortality.

As an example, Hudson notes that the fanatic classification of suicide terrorists implies that ideological and political rewards come before financial security, and view their actions as '*istishad*' (self-sacrifice for Allah) and martyrdom rather than

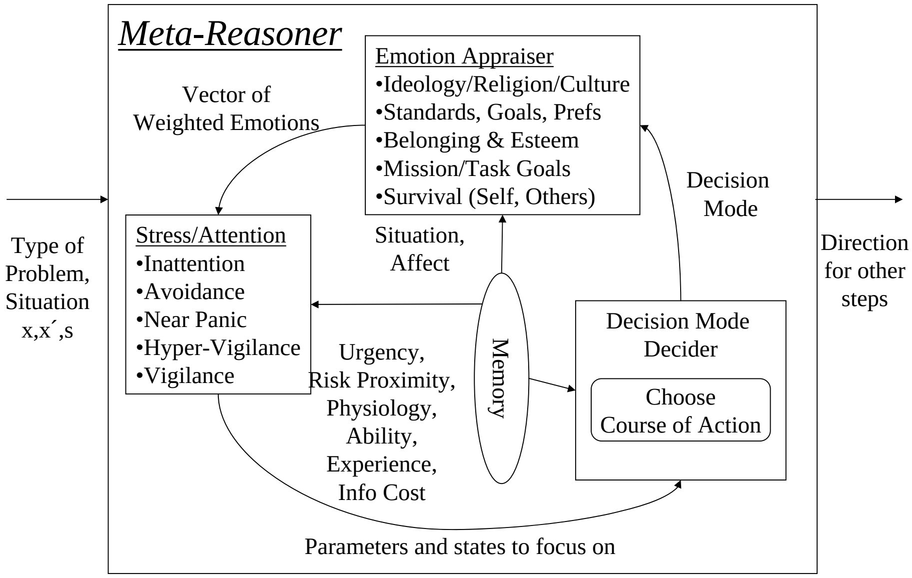


Figure 2.6: Meta-Reasoning Module [Silverman et al., 2001, pg. 5]

suicide [Hudson, 1999, pg. 31-4]. Such behavior could be programmed into the agent under this assumption. Accounting for how different or specific individuals adhere to this notion via a generalization of psychological, physiological, and external situations (e.g., economic and political status) has proved to be problematic in the past [Hudson, 1999, pg. 23, 30].

Assuming now that this model appropriately characterizes the key motives and resultant behaviors of terrorists, it is clear that different terrorist groups or individuals may require different behavioral functions in order to accurately account for the cultural, ethnic, gender, and situational differences. Weaver et al. [2001], in collaboration with Silverman et al. [2001], have developed a means to 'dial-up' an opponent such as a terrorist organization by leveraging the PMF approach. By accounting for the "(current) situation, organization, population, ideological/motivation, strategic, and tactical layers of their decision making," simulation setup of virtual opponents can be accomplished faster [Weaver et al., 2001, pg. 1].

Unfortunately, with the exception of painstakingly compiled historical data, limited information exists to accurately describe these layers. Further, this approach appears to limit the options of the simulated terrorist to actions that have occurred in the past (e.g., no surprise tactic of intentionally destroying a hijacked aircraft). However, given that an adversary constructed from and agent with artificial decision-making capabilities can be developed, the remaining area of interest is how the members of this project incorporated emotion into the decision-making process. This methodology is described by Johns and Silverman [2001].

Emotion, or in some cases the complete lack thereof, appears to play a large role in the decision-making capacity of terrorists and extremists, in general. Johns and Silverman attempt to develop a cognitive appraisal model that accompanies the meta-reasoning model depicted in Figure 2.6, with the goal of enabling agents to "...systematically reflect contextually relevant emotions and personality, and ..." study the affects upon their own decision-making behavior [Johns and Silverman, 2001, pg. 1]. Their work builds a relationship between an agent's emotions and the underlying "utility functions that drive decision theory" [Johns and Silverman, 2001, pg. 1]. This approach requires knowledge and understanding of the interaction between an agent's concerns and memory, both recent and long-term. Further, they divide concerns into "goals, standards, and preferences," which in the multi-objective context can be interpreted as objectives, costs [or constraints], and preferences [or value tradeoffs either within a single-dimensional value function or as indicated in a weight].

The decision-theoretic utility models that form the underlying structure of the model proposed should be scrutinized. For example, the conversion of emotion intensities (another, emotion-specific function that appears to be equivalent to a single dimensional value function) to utility of a course of action is provided by the following equation, $U\left(c\right)=\sum\limits_{d}\left(P_dE_x\right)$, where

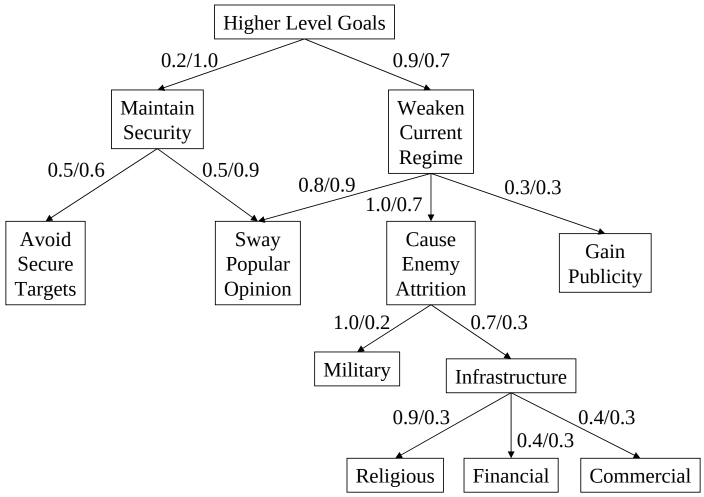


Figure 2.7: Goals for Terrorist A [Johns and Silverman, 2001, pg. 6]

> P_d is a measure of one of five personality dimensions (surgency, agreeableness, conscientiousness, openness to experience, and emotional stability); E_x is the "maximum intensity of emotion x (where x could be joy, anger, relief, etc.) over all possible concern effects times the perceived probability of this outcome actually occurring [Johns and Silverman, 2001, pg. 4-5].

Although explaining the specifics of these functions are beyond the scope of this research, a clearer signal that the underlying assumptions of the "pre-existing decision theory algorithms" claimed to be in use is shown in Figure 2.7. This hierarchy of goals is not comprised of mutually independent goals. Overall, the measures, functions, and weighting schemes do not appear to be of the classical form as seen in Keeney and Raiffa [1993].

2.3.11 Potential Improvements

Despite the potential theoretical and data acquisition issues associated with the overall agent-based framework, the goals identified and the realization that emotion must be incorporated within artificial decision-making systems in order to better represent and subsequently predict paradoxical behavior are of value. The goals shown in Figure 2.7 provide an excellent start at identifying the objectives that may

comprise a terrorist's value model. The component of emotion, however, could be included in either the shape of the single-dimensional value functions or, more appropriately, the preferences or intensity associated with the sub-goals or sub-objectives.

Due to the inherently non-cooperative nature of terrorist organizations, data collection will remain challenging but not impossible, as Krebs observed [Krebs, 2002, pg. 51]. Another possible solution to analyzing the problem of terrorism is to model *everyone else*, as they may be either cooperative or at least are not operating within secrecy and with the constant intent to deceive potential data acquisition efforts. One of the justifications for this given by Hudson is that "attempts to explain terrorism in purely psychological terms ignore the very real economic, political, and social factors that have always motivated radical activists..." [Hudson, 1999, pg. 23]. Hudson also notes that terrorism and political violence share underlying causes such as

> ...ethnic-, religious-, and ideological conflicts, poverty, modernization stresses, political inequities, lack of peaceful communications channels, traditions of violence, existence of revolutionary groups, governmental weakness and ineptness, erosions of confidence in regimes, and deep divisions within governing elites and leadership groups [Hudson, 1999, pg. 15].

Recent evidence that supports this theory is highlighted in an analysis of the groups responsible for the deadly and disruptive improvised explosive devices plaguing Coalition and Iraqi forces today. The conditions surrounding terrorists and their organizations, as well as measures to counter their activities may be explored within a simulation. The overall goal of such a simulation would be to investigate which of these factors can be addressed as to minimize the spread of terrorist support or the likelihood of recruitment.

2.3.12 Terrorist Models and 21st Century Warfare

Several authors, such as Johns and Silverman [2001], Silverman et al. [2001], and Weaver et al. [2001], are focusing on developing the capability to wargame

courses of action against various terrorist-based representations using agent-based simulations. Alternatively, these models could be used to predict what a terrorist might do, given a certain situation, providing insight into courses of action that might be sought by the terrorists themselves. This could improve prioritization of force protection efforts, such as securing the most likely target areas against attack. However, this approach may not provide insight into what the 'next great attack' will be, but only individual-individual or group-group agent interaction within a simulation. The simple act of constructing these methodologies has forced researchers to bring economic, psychological, and other genres of study together with focused background investigation of various terrorist organizations. This will inevitably act as a forcing function to better understand the enemy and possibly provide new means to defeat them even before they begin, for example, eliminate conditions that promote recruitment. This leads to another potential role of terrorist decision-making models, supporting the global war on terrorism.

Terrorism will likely not stop until all of the terrorists either change their ways or are destroyed. Extinction generally occurs in one of two ways: forced, such as over-hunting a specific species, or naturally, perhaps due to a sudden and drastic climate change to which the species cannot cope or adapt. *Forced*, in this context, requires the killing of all terrorists, assuming all terrorists could be properly identified; alternatively, *naturally* requires a change in the economic, political, social, and other appropriate climates that results in the destruction of the support structure and conditions that motivate and encourage terrorism. It has yet to be determined which of these are (1) more cost effective, including risk, (2) more effective in general, and (3) required in combination until the end of mankind.

As Harris suggested, researchers tend to model strange phenomena in order to improve our understanding. To better understand and eliminate terrorists and terrorist organizations, modeling is an appropriate means to facilitate understanding of their underlying psyche, motives, and overall goals. The unfortunate catch is that

efforts are inexorably biased by the analyst's own perceptions and ideologies, making it difficult to view decision-making from the target individual's perspective. As the theory of Maslow may not be entirely capable of describing terrorist decision-making, the notion of developing a cognitive appraisal (emotion) model acts as an incentive to deepen the study of a terrorist's behavioral characteristics as well as providing an opportunity to learn, predict, and directly or indirectly thwart this type of threat.

Upon review, several efforts have provided focused research and thought into these areas, but further work, particularly in the decision theoretic structures implemented within Johns and Silverman [2001], may be required for them to meet the most general assumptions of decision theory described in Keeney and Raiffa [1993]. Otherwise, current modeling efforts provide potential that it is possible to (1) gain insight into the underlying causes for motivating an individual to engage in these activities; (2) incorporate these concepts into agent-based simulations for study; and (3) use these simulations to evaluate various courses of action, ranging from close-combat operations, to application of other instruments of power or international diplomacy.

As a result of their simulation objectives to pit opposing sides against one another, these methodologies in general do not appear to focus on interactions that may occur within a terrorist network. Such interactions, often termed influence, pervade the social sciences literature and may hold a group together, break them apart, establish authority, and provide constraints upon or opportunities for social choices.

2.4 The Ebb and Flow of Influence

The seminal work of French described the then–current theory of social power and analyzed and addressed some of its limitations. In the course of his work, French defined "the basis of interpersonal power... as the more or less enduring relationship

between (two individuals) A and B which gives rise to power" [French, 1956, pg. 183]. He then described five bases for power: attraction, expert, reward, coercive, and legitimate [French, 1956, pg. 183-4]. In examining the impact of peer group influence upon opinion formation, Friedkin's interpretation of French's work was that "[French] first proposed that social influence was a finite distributed resource" [Friedkin and Cook, 1990, pg. 130]. Within the context of OR methodologies, Renfro postulated that influence was analogous to a commodity flowing through a (social) network [Renfro, 2001, pg. 80-1]. These and other works such as Freeman et al. [1991] substantiate the modeling the flow of influence as a commodity within a network model.

Measurement of influence in the context of social network analysis (SNA) is "based upon the importance of relationships among interacting (individuals)" [Wasserman and Faust, 1994, pg. 4]. Additionally, one of the underlying principles of SNA is that "...individuals view the network structural environment as providing opportunities for or constraints on individual action" [Wasserman and Faust, 1994, pg. 4]. This implies individuals take into account opinions of those socially close, or in positions of authority, for example, when faced with a decision point.

There are a variety of examples in SNA literature that investigate and attempt to measure this influence [cf. Frank and Yasumoto, 1988; Friedkin and Cook, 1990, among others]. A predominant concentration of research in this area deals with determining what conditions, both internal (via the network structure and connectedness of individuals) and external (via the outside influences or requirements for a group-supported decision), are required to bring a group of individuals to agreement upon a group decision.

Friedkin and Cook discuss social influence in the context of interpersonal relations within a network, and their subsequent role regarding the interpersonal influence required to enable the "...the process of (group) opinion formation" [1990, pg. 122]. This process utilizes network models to "...deal with the attainment of

collective agreements..., usually beginning with a network of fixed and discrepant opinions" [Friedkin and Cook, 1990, pg. 122]. Modeling the processes of "interpersonal negotiation" and the subsequent change in individual opinions form the "unique theoretical thrust of network models of social influence..." [Friedkin and Cook, 1990, pg. 122-3]. The resulting models essentially attempt to describe the dyadic interaction required to transform a network of individuals with discrepant opinions into a network where the individuals' opinions have coalesced, at least to some degree. Similar concepts in social network literature based upon an exchange of influence between individuals include contagion (of behavior) (Leenders [2002]) and diffusion (the rate of acceptance of innovative and possibly risky ideas or behavior) [Valente, 1996].

Just as there are many network model formulations within the OR domain, there exist numerous formulations and approaches within the study of social network modeling. A recent example, due to Amblard and Deffuant [2004], studied the propagation of extremist opinions throughout a variety of small-world networks. Their results suggest that "...a critical level of connectivity and some disorder in the network (is necessary) in order for extreme opinions to invade a population..." [Amblard and Deffuant, 2004, pg. 738].

However, this phenomenon is not necessarily confined to small-world networks. As Buchanan states the "infectious movement of desires and ideas from mind to mind is even the basis of a new theory of advertising known as permission marketing" [Buchanan, 2002, pg. 160-1]. Essentially, this connotes the flow of influence propagating through a general populous, which may not necessarily be a small-world network in the classical sense. This is an important point because not all organizations may naturally evolve as small-world networks. However, influence will inevitably flow regardless of the underlying network structure [Renfro, 2001].

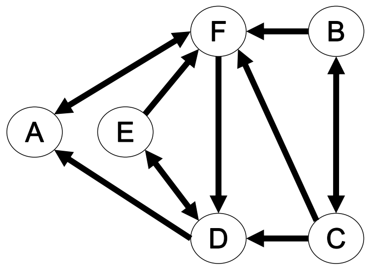
Beginning with French's influential work, it is clear that the social science research and theory liken the interaction between two individuals or groups to that of

a commodity that flows between them. Operations Researchers and Social Scientists generally apply network models differently, a key difference being that social science models tend to be descriptive, while OR models tend to be both descriptive and prescriptive, where appropriate. Descriptive models, in general, attempt to describe how a process or system works via underlying relationships and behaviors. The focus of prescriptive models is improved decision making by attempting to describe the best or optimal solution of a given system [Clemen, 1996, pg. 14]. Oftentimes, the process of obtaining a prescriptive model requires an understanding of the underlying processes or systems inherent to the decision problem and therefore results in a descriptive model as a byproduct. The next section discusses a few of the seminal, descriptive measures used within the SNA literature to ascertain individuals of interest. Where appropriate, critiques and potential areas of theoretical improvement of these measures are offered.

2.4.1 Katz Status Index

In an attempt to improve upon the prevalent status measures, Katz, in the context of a popularity contest, based individual status not only upon how many people choose the most popular individual but also accounting for who is doing the choosing. Katz suggested that this measure may also be "...used to study influence, transmission of information, etc." [Katz, 1953, pg. 39].

Katz notes that the column sums of $\mathbf{X}$ pertain to the number of people that choose that individual; this form of 'in-degree' centrality was the primary means for assessing status during the time of his research. Further, noting that the elements of the powers of the sociomatrix, given by $\mathbf{X}^p$, provide the number of directed walks of length p from i to j, he noted that this equates to the indirect p-step ($p > 1$) choices of a given individual by the group [Katz, 1953, pg. 40] [cf. Wasserman and Faust, 1994, pg. 160-1]. All possible walks are accounted for by raising the sociomatrix to the power of infinity. An additional assumption that longer walks were less effective



$$\mathbf{C} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Figure 2.8: Katz Choice Matrix [Katz, 1953, pg. 40]

than shorter ones required an attenuation factor, $0 \leq \alpha \leq 1$. Accepting these constructs, Katz' objective was to find the column sums of the matrix

$$\mathbf{T} = \alpha\mathbf{C} + \alpha^2\mathbf{C}^2 + \alpha^3\mathbf{C}^3 + \cdots + \alpha^k\mathbf{C}^k + \cdots \tag{2.1}$$

Given the computational limitations of the early 1950s, Katz cleverly sought to take advantage of the geometric series, shown in Equation 2.2, avoiding the computation of matrix powers:

$$\sum_{k=1}^{\infty} r^k = \frac{r}{1-r} \ , \quad r < 1 \ . \tag{2.2}$$

Consequently, substituting $r = \alpha\mathbf{C}$ and applying Equation 2.2 to Equation 2.1 yields

$$\mathbf{T} = (\mathbf{I} - \alpha\mathbf{C})^{-1} - \mathbf{I}. \tag{2.3}$$

Considering that the conventional status index divides the column sums by $(n-1)$, the column sums of $\mathbf{T}$ are divided by the m value that accommodates the new construct's underlying technique [Katz, 1953, pg. 42]. Figure 2.8 illustrates the choice matrix and accompanying digraph.

$$m \cong (n-1)!\alpha^{(n-1)}e^{1/\alpha} \tag{2.4}$$

Given this information, the original status vector (one element for each of the six actors) is $\mathbf{s} = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.6 & 0.2 & 0.8 \end{bmatrix}$; essentially, those actors with high in-degree (that is, actors F, D, and A in descending order) dominate with regards to status. Alternatively, the status vector using Katz's measure, with a multiplier of $\alpha = 0.5$, is $\mathbf{s} = \begin{bmatrix} 0.47 & 0.04 & 0.04 & 0.41 & 0.22 & 0.45 \end{bmatrix}$. Using Katz's approach, actor A scores higher than actor F, albeit slightly. Despite the relatively low in-degree of actor A, his status is highest because both of the actors with the highest in-degree (actors F and D) choose actor A. Directly stated, "being chosen by a popular individual should add more to one's popularity" [Bonacich and Lloyd, 2001, pg. 192]. The change in status for actors B, C and E from being equivalent to E differing from B and C is accounted for in a similar fashion [Katz, 1953, pg. 42]. Interestingly, Katz does not offer an interpretation of the elements of the $\mathbf{T}$ matrix.

A few points of contention exist: the characteristics of the flow captured or assumed by the calculations; the potential length of the paths implicitly accounted for within the measure's calculations; and, the arbitrary choice of the attenuation factor. Assuming that this methodology can indeed be applied to the transmission of information, the matrix powers ($p > 2$) actually capture a variety of walks that may not necessarily be conducive to operations security. Deo offers a more precise definition of the content of the powers of the sociomatrix, which is summarized in the following theorem.

Theorem 1. *The (i, j)th entry in $\mathbf{X}^p$ equals the number of different, directed edge sequences of r edges from the ith vertex to the jth. These sequences fall into three categories:*

1. *Directed paths from i to j: those directed edge sequences in which no vertex is traversed more than once;*
2. *Directed walks from i to j: those directed edge sequences in which a vertex may be traversed more than once, but no edge is traversed more than once; and,*

3. *Those directed edge sequences in which an edge may also be traversed more than once* [Deo, 1974, pg. 222].

Leenders previously pointed out that the information contained within the powers of the sociomatrix is often misperceived, depending upon the operative definition of 'walk' [Leenders, 2002, pg. 32]. Additionally, the second and third categories of information flow are likely contrary to the security goals of a non-cooperative network. For example, with $p = 4$, a possible walk of four between A to D will include A-F-A-F-D. If the network of interest is trying to maintain secretive communications, as in the case of Al Qaeda, the banter between A and F may be unlikely [Post, 2005, pg. 39-48].

In addition, the potential length of walks measured goes to infinity. Again, this would involve an infinite amount of communication between the individuals, which would likely be counter to their security objectives. This suggests that a more direct, path-based approach, limited to the length of the longest path given n actors–or $(n-1)$–would be more appropriate. In fact, although the approximation of the denominator m (from Equation 2.4) is based upon an infinite series, the elements of this series reduce to zero when considering powers beyond $(n-1)$. However, using $(\alpha = 0.5)$ and the choice matrix discussed by Katz, the summation due to $(p \geq 7)$ contributes a significant amount to the overall measure.

Finally, the arbitrary 'attenuation' factor has been highlighted by previous works [Clark, 2005; Borgatti and Everett, 2006]. Although the value for α is likened to the 'attenuation' of a signal or influence as a function of distance traveled, it is simply required for the series to converge, thereby providing a result. For his measure, Katz suggests "that reasonable, general-purpose values of α^{-1} are those between the largest [eigenvalue of $\mathbf{X}$] and about twice that [value]" [Katz, 1953, pg. 42]. Using the same example discussed by Katz, the largest eigenvalue of $\mathbf{X}$ (which corresponds to Figure 2.8) is 1.68; this implies that $0.298 < \alpha < 0.595$. Hence, the assumption space for 'attenuation' within which the analyst can work is restricted

from the onset. Additionally, even within this recommended range, the most 'central' actor is ultimately a function of α.

Note also that this approach assumes that each link (or the strength of a link) is identical to all others. Hence, computational methods that (1) do not rely upon an arbitrary input merely for convergence and (2) account for weighted links are potentially of interest. Two examples of such measures include the clique identification approach by Hubbell [1965] and information centrality developed by Stephenson and Zelen [1989].

2.4.2 Mechanics of Clique Identification

Hubbell's objective was to also use the concepts of status and choice in determining cliques within a network. His measure is related to Katz's in that it accounts for the status of the chooser, but also incorporates "the strength at which he chooses" [Hubbell, 1965, pg. 382]. Instead of an arbitrary attenuation factor, Hubbell requires that the value w_{ij} be specified–which may be negative, zero, or positive–for each pair of actors [Hubbell, 1965, pg. 378]. As noted in Katz's measure, the matrix of these weights ($\mathbf{W}$) is raised to powers to account for indirect paths of influence. Consequently, w_{ij}^{p} corresponds to "the total strength of j's influence upon i at the p^{th} remove" [Hubbell, 1965, pg. 379]. Relying again upon a geometric series, Hubbell defines the index of association (m_{ij}) to "discriminate intra-clique bonds from inter-clique bonds" [Hubbell, 1965, pg. 379].

$$y_{ij} = \delta_{ij} + w_{ij} + w_{ij}^{(2)} + w_{ij}^{(3)} + \ldots \tag{2.5}$$

$$m_{ij} = m_{ji} = \min\left(y_{ij}, y_{ji}\right) \tag{2.6}$$

An additional extension posed by Hubbell permits the incorporation of exogenous variables (e_i) specific to each actor into the status measure. Unfortunately, although constraints upon w_{ij} are specified in order for a solution to exist, the development

and underlying theory behind all of these values are vague at best. In addition, the values of m_{ij} are compared against a relatively arbitrary threshold in order to discern pairs within a clique [Hubbell, 1965, pg. 379]. Nonetheless, the notion of strengths of relationships through the weighted values and the incorporation of information other than network topology was a substantial improvement over previous measures. Information centrality, discussed next, offered further improvements.

2.4.3 Information Centrality

Stephenson and Zelen developed a centrality measure based upon the amount of information (potentially) flowing through "all paths between [all] pairs of points" [Stephenson and Zelen, 1989, pg. 2]. This approach is attractive in that (1) it accounts for the geodesic as well as the non-geodesic communication or flow of influence through a network and (2) it easily permits the use of weighted graphs. Further, the measure is not based upon powers of the sociomatrix, thereby characterizing the potential flows of influence that are more likely within the realm of non-cooperative network behavior. The author's unique approach avoids the necessity to explicitly enumerate all paths; however, this ultimately restricts application of this measure to symmetric graphs [Stephenson and Zelen, 1989, pg. 4].

The underlying motivation for developing information centrality was to address the fact that betweenness and closeness measures essentially neglected "...measuring communication occurring along reachable, non-geodetic pathways" that may be leveraged by a particular organization [Stephenson and Zelen, 1989, pg. 3]. The authors had considered that communication "...may be intentionally channeled through many intermediaries in order to hide or shield information in a way not captured by geodesic paths" [Stephenson and Zelen, 1989, pg. 3]. This aspect alone suggests that it may be an attractive option for analysis of non-cooperative networks.

Using the length of the path as a distance, the information contained on a path from i to j is defined as the reciprocal of this distance. The authors based

their approach on "theories of the statistical design of experiments and estimation" [Stephenson and Zelen, 1989, pg. 28]. Ultimately, Stephenson and Zelen posit that this approach more effectively captures "subtle network infrastructures in complex situations" [Stephenson and Zelen, 1989, pg. 27]. Such complex situations are likely those encountered by members in a non-cooperative organization attempting to limit exposure to their communications. A final genre of centrality measures that also avoids the common assumption of 'efficient communications' is based upon the eigenvectors and eigenvalues of the sociomatrix.

2.4.4 Eigenvector Centrality

The underlying premise of eigenvector centrality is summarized by Bonacich and Lloyd.

> Being chosen by a popular individual should add more to one's popularity. Being nominated as powerful by someone seen by others as powerful should contribute more to one's perceived power. Having power over someone who in turn has power over others makes one more powerful [Bonacich and Lloyd, 2001, pg. 192].

Mathematically, given the entries in the adjacency matrix, denoted (a_{ij}) for this measure, implies that i contributes to j's status, and x_i denotes the status of individual i, this concept is shown in Equation 2.7 [Bonacich and Lloyd, 2001, pg. 192-3].

$$x_i = a_{1i}x_1 + a_{2i}x_2 + \ldots + a_{ni}x_n \tag{2.7}$$

In order to determine solutions to this system, a generalized form is shown in Equation 2.8, with a specified scalar λ.

$$\lambda x_i = a_{1i}x_1 + a_{2i}x_2 + \ldots + a_{ni}x_n \tag{2.8}$$

In matrix notation, this is denoted by Equation 2.9, which is commonly known as the eigenvalue problem. Note that for this particular measure, $\mathbf{X}$ is an $n \times n$ matrix

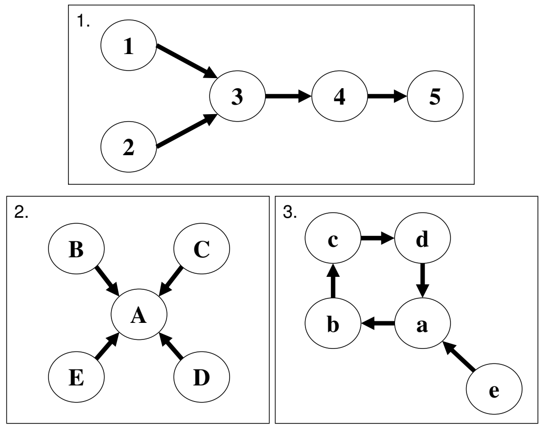


Figure 2.9: Hypothetical Network [Bonacich and Lloyd, 2001, pg. 192]

with columns comprised of the eigenvectors of the adjacency matrix ($\mathbf{A}$) and λ is an $n \times n$ diagonal matrix of the eigenvalues of $\mathbf{A}$ [Bonacich and Lloyd, 2001, pg. 193].

$$\mathbf{A}^{\mathrm{T}}\mathbf{X} = \mathbf{X}\lambda \tag{2.9}$$

As expected, network structure plays an important role in the results of this analysis method. However, there are unique cases where the numerical results may not capture the more intuitive understanding of centrality. For example, all actors within each of the hypothetical, directed networks shown in Figure 2.9 have zero status due to "...positions that receive no choices have no status and contribute nothing to any other position's status" [Bonacich and Lloyd, 2001, pg. 139].

To get around this conceptual and mathematical issue, Bonacich and Lloyd proposed "α-centrality" that provides every individual some level of status, independent of existing or non-existent connections to others [Bonacich and Lloyd, 2001, pg. 193]. With the vector of exogenous sources of status ($\mathbf{e}$) and a parameter reflecting the "...relative importance of endogenous versus exogenous factors in the determination of centrality" (α), the matrix solution for status is given by

$$\mathbf{x} = \alpha\mathbf{A}^{\mathrm{T}}\mathbf{x} + \mathbf{e} \Rightarrow \mathbf{x} = \left(\mathbf{I} - \alpha\mathbf{A}^{\mathrm{T}}\right)^{-1}\mathbf{e}. \tag{2.10}$$

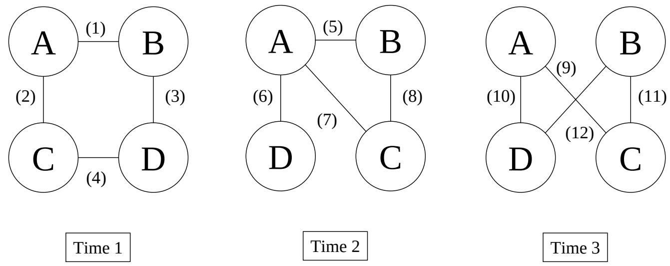


Figure 2.10: Notional Hypergraph [Bonacich et al., 2004, pg. 190]

Note that Katz's model is similar to this approach, and differs by a constant of one [Bonacich and Lloyd, 2001, pg. 194]. Although the theoretical development, range, and magnitude (other than being a vector of ones in their example) of $\mathbf{e}$ are not discussed, the new approach both permits analysis of asymmetric relationships and is equivalent to the original formulation as α approaches $\lambda_{\max}^{-1}$ [Bonacich and Lloyd, 2001, pg. 196-7]. Clearly the development of the theoretical nature of the exogenous vector offers a target of opportunity; for example, an extension of the discriminate analysis technique used by Clark [2005] to ascertain actor position may provide a useful first step.

A recent extension to this concept facilitates the determination of centrality for hypergraphs and hyperedges, which account for the effects of multiple dimensions within a relationship (e.g., time, place, and group membership) [Bonacich et al., 2004, pg. 192] [cf. Seidman, 1981, which addresses some methods to deal with social structures via hypergraphs]. For example, consider the set of four actors with relationships or interactions captured at three different time periods (see Figure 2.10). Representation of all three graphs simultaneously may be accomplished via an incidence matrix where "each edge is represented by a row and each vertex by a column" [Bonacich et al., 2004, pg. 193]. Note that this storage format also prohibits the analysis of directed and/or weighted graphs. From Figure 2.10, the number in parenthesis (i) corresponds to the ith row in the incidence matrix shown below. The first four columns correspond to the actors A, B, C, D and the last three columns

correspond to the time periods 1, 2, 3. As an example, the first row of matrix $\mathbf{E}$ indicates that actors A and B are connected in time period 1:

$$\mathbf{E} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.11}$$

For the complete set of results, the reader is referred to [Bonacich et al., 2004, pg. 195]. According to the centrality values and Bonacich et al., one may conclude that actor A (C) is the most (least) central; time period 2 the most central among all three time periods; and, hyperedge 7 is the most central [Bonacich et al., 2004, pg. 195]. Unfortunately, the interpretation of the centrality of a network within a set of networks remains unclear. Clark used multidimensional centrality between networks as weights, after normalization, as a proxy for the importance of a given contextual network [Clark, 2005, pg. 3-29].

This technique, like many others, relies solely upon the network topology present within a given relationship context as opposed to the importance of the context from a cultural perspective of the actors within that network. For example, given a set of actors, suppose that known relations within a given context are fewer in number than another context considered in the analysis. It is hypothesized (and

therefore an issue that must be addressed in the underlying theory/methodology) that the lack of data (for one reason or another) will affect the multidimensional centrality score for that contextual network. Suppose further that this smaller network is a context that plays a significant role in the solidarity of the group, perhaps much more than any of the other contexts. Using multidimensional centrality of each network context, therefore, has the potential to underweight the key relationships that exist among multiple contexts. This leads into the next concept of interest within this research–multiplexity.

2.4.5 Multiplexity and Layered Networks

Given a set of actors, when more than one relationship or context of interaction is studied the analysis is considered multiplex [Monge and Contractor, 2003, pg. 35]. This term, like many other concepts in SNA, appears to be borrowed from communications theory, which defines multiplex as combining multiple signals into one to facilitate transmission, in such a way that they can later be separated as required [DOD, 2005, pg. 354]. Consequently, communication and interaction between two individuals will generally transmit through several different contexts simultaneously. As Haythornthwaite noted, "we operate in a multiplex world, maintaining multiple roles and relations with others, sustained through a variety of media" [Haythornthwaite, 1999].

Although interaction between two individuals naturally involves these multiple relations (for example, family, friend, co-worker, fellow student, any combination thereof, and so forth), surprisingly few articles actually incorporate multiplex relations within their analysis. Such lack of previous studies may be attributed to the complexity encountered when dealing with multiplex networks. Interestingly, Wasserman and Faust recommend that commonly used centrality and prestige measures be calculated for each relation separately and recommend against aggregating the relations into one sociomatrix [Wasserman and Faust, 1994, pg. 219]. Although

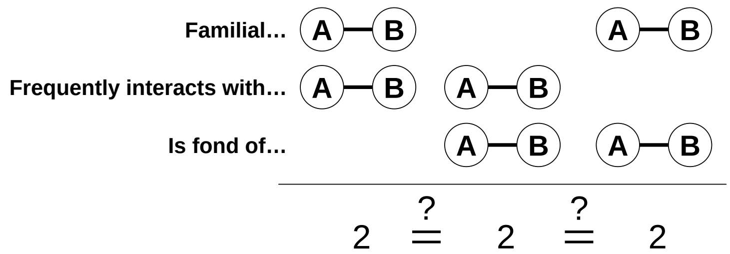


Figure 2.11: Improper Multiplex Aggregation

rationale for this is not provided, the answer is likely the loss of information incurred when merely combining occurrences of links among relations.

For example, consider the three possible instances within which two individuals can share relations in two out of three contexts (Figure 2.11). Assuming that the contexts were 'familial,' 'frequently interacts with,' and 'is fond of,' one could posit examples where all of these possibilities would yield different strengths of relationships. However, a simple summation of dichotomous occurrences results in identical 'strengths' and is therefore likely insufficient to capture or infer the strength of a relationship based upon multiplex data.

Nonetheless, when two people interact, regardless of the value of the relationship's strength or a means to quantify it, it is assumed that both actors are cognizant of the underlying contexts that prevail and make their relationship either strong or tenuous. This may imply that social network measures, applied to each of the networks or contexts independently, will fail to capture the combined effect due to the multiplexity inherent within the relationships. This suggests that, prior to determination of centrality, prestige, and so forth, an aggregation of contexts would be analytically prudent. One potential means could comprise a weighted function, based upon how the actors internal to the network of interest place importance upon each context. A notional example of this is illustrated in Figure 2.12. Of course, one could ask the question "Is a familial link equivalent in strength to a tie that shares both the bonds of fondness and frequent interaction, since the weighted sum

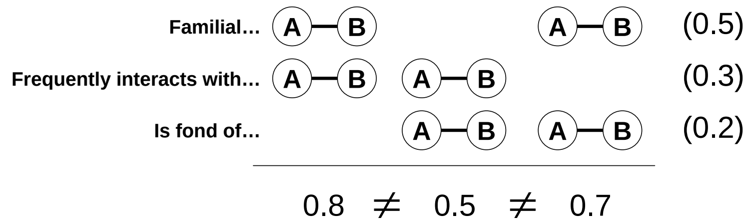


Figure 2.12: Potential Aggregation Scheme

in either case equals 0.5?" The most likely answer is 'it depends,' and is therefore an underlying question within this research.

In order to begin an understanding of weighting schemes, their limitations, and their strengths, a review of decision analytic weighting techniques is discussed.

2.5 Weighting and Decision Analysis

> Decisions are a "ubiquitous activity," that involves a "commitment of resources. Resources need not be financial, natural or even material" [LaValle, 1978, pg. 3].

In today's complex and interdependent world, the commitment of resources LaValle cites is not necessarily limited to implementing a decision. Additionally, it often entails up-front data gathering, decision maker (DM) and stakeholder involvement, and analysis required of thoughtful and conscientious decision making. Within the multi-criteria decision making (MCDM) framework, this section focuses upon one such up-front activity–the construction of weights, also referred to as scaling constants, for additive multi-attribute value functions.

With a few assumptions regarding the decision environment, the nature of the decision problem, and the expense of up-front costs, a conceptual extension to a variety of weighting schemes is possible. In the field of multi-attribute value theory, or more generally, multi-objective decision analysis, there exist a number of elicitation techniques to weight attributes within a value model. Each technique has its strengths and weaknesses; however, all are subject to expenses such as time,

money, and judgment errors and biases due to the requisite decision-maker(s) and analyst(s) interaction and involvement. One of the focal points of this research seeks to economize upon this aspect of decision analysis. Beginning with a brief review of the MCDM approach and the circumstances under which the proposed methodology shows promise, the evolution and sometimes conflicting paradigms of weighting are discussed. Extending a few of these techniques to permit a dynamic realization of weighting vectors is proposed in Chapter VI; this concept proves useful in not only providing the theoretical underpinnings of weighting the contextual layers of a social network, but also permitting dynamic weighting of alternatives, of infrastructure network layers, and other MCDM-oriented applications.

2.5.1 Multi-Criteria Decision Making

MCDM provides a systematic approach for thinking about and structuring objectives in the context of a given decision problem. Decisions are made to meet an overall objective where, in general, "an objective generally indicates the direction in which we should strive to do better...," in the context of a given decision problem [Keeney and Raiffa, 1993, pg. 32-4]. Unfortunately, within the context of today's multifaceted decision environment, the overall objective may almost certainly be accompanied by conflicting sub-objectives. This environment comprises the inherent nature of complex decisions in that trade-offs must be made when selecting a 'best' course of action or alternative [Keeney and Raiffa, 1993, pg. 15].

The prescriptive nature of MCDM forces decision makers to think critically about the various problem dimensions, breaking objectives into sub-objectives, and continuing this process until each sub-objective may be measured by an attribute. An attribute "indicates the degree to which alternative policies meet this objective" [Keeney and Raiffa, 1993, pg. 32]. The desirable properties are summarized in Table 2.10 [Keeney and Raiffa, 1993].

Table 2.10: Attribute Properties [Keeney and Raiffa, 1993, pg. 50-3]

Complete	Attributes cover all important aspects (sub-objectives) of the problem
Operational	Useful and meaningful to the decision maker; facilitates decision making
Decomposable	The nature of the attributes allows simplification of the evaluation process by breaking it into parts
Non-redundant	Limits double counting
Minimal	Keeps the problem dimension small as possible

The MCDM process seeks to "systematically think about ranking a set of consequences when each consequence is described in terms of performance values on many attributes" [Keeney and Raiffa, 1993, pg. 28]. For the purposes of this research, it is assumed that this task is achieved via a multi-attribute, additive value model. Additive models, as opposed to multiplicative models, are not only easier to understand and analyze, but also perform well despite some of the theoretical criticisms. For example, Stewart devised a simulation experiment to test the robustness of additive models and their ability to "reproduce the 'ideal' preference ordering of the alternatives" [Stewart, 1996, pg. 305]. Stewart's findings indicate that, with some care exercised by the decision analyst, additive value models perform well despite violations of underlying assumptions such as non-linearities in the single dimensional value functions, additive independence, and inadvertent omission of a small portion of the model criteria [Stewart, 1996, pg. 308].

2.5.2 Additive Value Model

A value model offers a means to quantitatively measure a decision maker's preferences. Assessments of this nature trace back to the psychophysical study of judgments regarding subjective phenomena (e.g., loudness, pitch, and brightness) [von Winterfeldt and Edwards, 1986, pg. 209]. Decision science attempts to leverage this field, enabling the quantitative measurement of the subjective phenomena–preference [von Winterfeldt and Edwards, 1986, pg. 209].

Let the vector, $\mathbf{x}$, represent the characterization of a given alternative (or act), a, within an n-dimensional consequence space such that $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Each dimension corresponds to a measurable attribute. Therefore, the characterization is achieved through evaluating the alternative with respect to each attribute. A single-dimensional value function (SDVF) that evaluates the given alternative with respect to ith attribute provides the realization of x_i, mathematically shown as $\mathrm{X}_i(a) \equiv x_i, \ \forall\ i$. In general, a value function, connoting decision making under certainty, "associates a real number $v(\mathbf{x})$ to each point $\mathbf{x}$ in an evaluation space ..., [that represents] the decision maker's preference structure provided that" the function output can discern between indifferent and preferred alternatives [Keeney and Raiffa, 1993, pg. 80]. Given two alternatives denoted x' and x'', Keeney and Raiffa highlight these two requirements mathematically in Equations 2.12 and 2.13, respectively.

$$x' \sim x'' \Leftrightarrow v(x') = v(x'') \tag{2.12}$$

$$x' \succ x'' \Leftrightarrow v(x') > v(x'') \tag{2.13}$$

These mappings between preference and the value function output, particularly from Equation 2.12, play a key role in the weight elicitation method proposed by Keeney and Raiffa [1993], discussed later in Section 2.5.5.2.

Although there are a number of underlying forms for value functions, the model assumed for this effort is the simple additive value model (cf. [Keeney and Raiffa, 1993, pg. 81] and [von Winterfeldt and Edwards, 1986, pg. 276]). As mentioned by Keeney and Raiffa [1993], verification of the requisite independence assumptions for multiplicative models can be difficult with even relatively small models of five or more attributes. Keeney also notes that "...if any of the independence conditions are not appropriate, it is an indication that an objective in addition to those articulated for the problem is relevant" [Keeney, 1988, pg. 153]. Therefore, another alternative to implementation of the multiplicative model could include revisiting the

objectives hierarchy and determining whether or not an additional objective and its corresponding attribute should be incorporated within the model. Nonetheless, the dynamic weighting approaches developed later could be applied to either construct.

According to Keeney and Raiffa [1993, pg. 91], in order for the additive preference structure to hold, the model must be in (or transformable to) the following form,

$$v(x_1, x_2, \ldots, x_n) = v_{X_1}(x_1) + v_{X_2}(x_2) + \cdots + v_{X_n}(x_n). \tag{2.14}$$

From Equation 2.14, X_i, $(i = 1, \ldots, n)$ indicates the ith attribute, whereas x_i, $(i = 1, \ldots, n)$ indicates the score of a given alternative with respect to the ith attribute.

Definition 9. *The attributes* $X_1, \ldots, X_n$ $(n \geq 3)$ *are* mutually preferentially independent *if every subset* Y *of these attributes is preferentially independent of its complementary set of evaluators* [Keeney and Raiffa, 1993, pg. 111].

Theorem 2. *Given attributes* $X_1, \ldots, X_n \, (n \geq 3)$*, an additive value function of the form*

$$v(\mathbf{x}) = \sum_{i=1}^{n} v_i(x_i), \tag{2.15}$$

exists if and only if the attributes are mutually preferentially independent [Keeney and Raiffa, 1993]. *Note that for Equation 2.15, the subscripts* X_i *are simply replaced with* i *for notational convenience.*

The final component of the value function involves the weights, often referred to as scaling constants. Given an additive value function of the form shown in equation 2.15, Keeney and Raiffa note that, for purposes of convenience, the overall score, $v(\mathbf{x})$, as well as the single-attribute [or single-dimensional] value functions, $v_i(x_i)$, should be scaled [Keeney and Raiffa, 1993, pg. 117]. This results in the form,

$$v(\mathbf{x}) = \sum_{i=1}^{n} w_i v_i(x_i), \tag{2.16}$$

which is also an additive value function equivalent to Equation 2.15, assuming that the scaling for both the single dimensional value functions and the weights is consistent [Keeney and Raiffa, 1993, pg. 116-7].

Note that Equation 2.16 also assumes that the weights are normalized such that $\sum_{i=1}^{n} w_i = 1$. It is important to understand the underlying rationale for normalizing the weights, and their interpretation, prior to extending the various weighting methodologies; proposed extensions are discussed in Chapter VI.

2.5.3 Normalization of the Weights

Similar to the reason observed by Keeney and Raiffa, Lootsma suggests that weight normalization within this modeling approach provides "a uniform scale to judge the alternatives under the respective criteria, ...(and to) easily quantify the gradations of the relative importance of the criteria" [Lootsma, 1999, pg. 36]. Essentially, if the single-dimensional value functions are all scaled consistently (e.g., all range from 0 to 1, 0 to 10, or 0 to 100, as suggested by the decision environment), normalizing the weights will result in an overall value function score that is within the same range. This facilitates interpretation of the overall scores yet does not change the end result with respect to the preference order of the scored alternatives.

For example, suppose a three-attribute value model, comprised of SDVFs, $v_i(x_i) \in [0,1]$ and $i = 1,2,3$, were developed and weights elicited. In Table 2.11, the *Raw* weights are the original numbers elicited; the corresponding *Normalized* weights carry the same effect but now sum to 1. The scores in Table 2.12, based upon each set of weights, results in the same rank order by preference. However, using the normalized scores results in an overall score $v(\mathbf{x}) \in [0,1]$, which is the same range as each of the single-dimensional value functions. This is referred to as the consistency condition [Keeney and Raiffa, 1993, pg. 271].

Note that unless the value function is a measurable value function, the differences between alternatives' scores indicate ordinal preference ranking only. Assuming

Table 2.11: Weight Normalization

	w_1	w_2	w_3
Raw	81	162	15
Normalized	0.31	0.63	0.06

Table 2.12: Consequence of Normalized Weights

	$v_1(x_1)$	$v_2(x_2)$	$v_3(x_3)$	Score (Normalized)	Score (Raw)
a_1	0.50	0.80	0.10	0.67	171.6
a_2	0.60	0.70	0.50	0.66	169.5
a_3	0.10	0.40	0.30	0.30	77.4

the example in Table 2.12 is not a measurable value function, it can be stated that the decision maker prefers a_3 over a_2, but it cannot be stated that the value of a_3 is over twice the value of a_2 [cf. Dyer and Sarin, 1979; Kirkwood, 1997, pg. 241-4].

Normalization of the weights subsequently limits the weight-sets, $\mathbf{w}$, to lie within a bounded polyhedral set such that $W = \{\mathbf{w} : \mathbf{0} \leq \mathbf{w_i} \leq \mathbf{1}, \forall i\}$; this was also observed, and leveraged for their studies, by Wolters and Mareschal [1995, pg. 283] and Ma et al. [2001, pg. 67]. The weight space for two- and three-dimensional value functions are shown in Figure 2.13 and Figure 2.14 respectively. Graphically, the two-dimensional weight space is a line, and the three-dimensional weight space is the triangular plane. In general, the bounded polyhedral is formed in any n-dimensional ($n \geq 2$) space. Theoretically, and as a consequence of the structure of W, any combination of weights within the set W is possible, although some are not as likely, such as having the weight of one attribute equal to zero.

2.5.4 Interpretation of Normalized Weights

When interpreting the normalized weights for each attribute within a value model, Keeney and Raiffa emphasize that the values derived for the weights do not indicate the attributes' relative importance [Keeney and Raiffa, 1993, pg. 271-2]. Interestingly, this common misinterpretation underlies a few of the weighting ap-

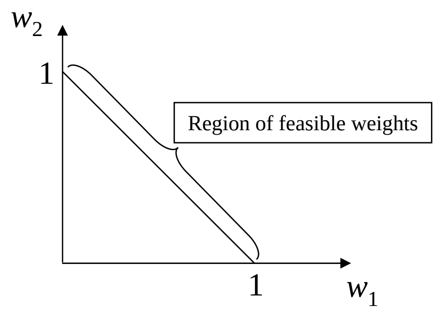


Figure 2.13: Weight Space (2-D)

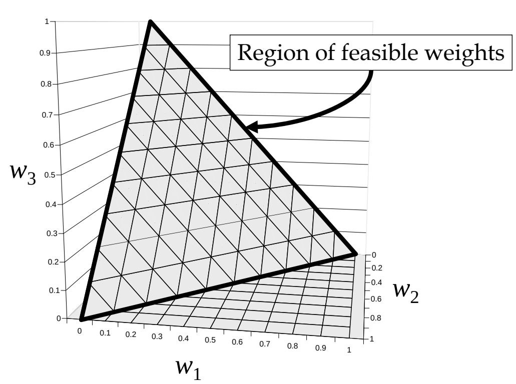


Figure 2.14: Weight Space (3-D)

proaches either in use or provided in popular decision analysis texts, particularly the direct weighting schemes discussed later in Section 2.5.5.1. Several issues contribute to this quandary.

Theoretically, although Keeney and Raiffa appear to develop the construct of weights from an economic perspective (i.e., the marginal rate of substitution), the substitution effect or relationship may no longer be relevant in the context of an additive value function [Keeney and Raiffa, 1993, pg. 74-77]. Schenkerman, in his discussion of the general abuse and misunderstanding of weights, concludes that the weights are "estimates of marginal preferences..." that capture decision-maker tradeoffs [Schenkerman, 1991, pg. 371-2]. However, Schenkerman also notes that constant rates of substitution implies all of the single-dimensional value functions are linear–a circumstance that cannot be guaranteed for all decision models and is easily accounted for by exponential, piece-wise linear, and other non-linear single-dimensional value functions discussed in the literature [Schenkerman, 1991, pg. 372] [For examples, see Kirkwood, 1997, pg. 64-68].

In order to understand the nuances of weighting and the potential issues that may arise due to dynamic extensions of the weighting methodologies, a few of the methods prevalent within literature are discussed.

2.5.5 Weighting Methodologies

As shown in Table 2.13, von Winterfeldt and Edwards provide a cogent taxonomy of techniques used to construct attribute weights. These techniques have spawned a variety of specialized procedures, some of which will be discussed; however, those methods developed since 1986 appear to be a derivation of one or more of the techniques shown in Table 2.13.

2.5.5.1 Numerical Estimation

Table 2.13: Weighting Taxonomy [von Winterfeldt and Edwards, 1986, pg. 274]

Numerical Estimation Methods	Indifference Methods
Ranking Direct Rating Ratio Estimation Swing Weighting	Cross-attribute Indifference Cross-attribute Strength of Preference

von Winterfeldt and Edwards state that *ranking* and *direct rating* are simplified variations of *ratio estimation*. Ranking generally involves the ordering of the attributes from most to least important. Rating is often accomplished by the distribution of a limited number of points to each of the attributes to capture relative importance [von Winterfeldt and Edwards, 1986, pg. 284]. The family of simple multi-attribute rating techniques (SMART, SMARTS, and SMARTER) comprise the ratio estimation, swing weighting, and ranking methodologies. Although, the numerical estimation techniques are generally the easiest to implement, they are not without their disadvantages. For example, numerical estimation techniques "explicitly involves the notion of attribute importance," criticized by Keeney and Raiffa [1993, pg. 271-2] and often fail to capture effects due to subsequent changes within the attribute ranges von Winterfeldt and Edwards [1986, pg. 285]. Another disadvantage, specifically associated with the ranking and rating techniques, involves the type of numbers involved and the calculations performed on them.

A variety of ranking procedures are described by von Winterfeldt and Edwards [1986] and Stillwell and Edwards [1979]. Those discussed in detail below include: the *rank reciprocal rule*, *rank sum*, *rank exponent*, and *decision rule ranking*. For each of these techniques, R_i is the rank for attribute i and the model accounts for a total of j attributes. Decision rule ranking is a two-fold technique that elicits both an ordering of attributes as well as an estimate for the weight of the most important attribute. The ranking method chosen is the one that "...most closely approximates the weight elicited for the first dimension" [Stillwell and Edwards, 1979, pg. 11]. All of these elicitation approaches, shown in Table 2.14, are among the most straight-

Table 2.14: Ranking & Rating Methods [von Winterfeldt and Edwards, 1986, pg. 284]

Method	Formula
Rank Reciprocal Rule	$w_i = \frac{1/R_i}{\sum_j 1/R_j}$
Rank Sum	$w_i = (n+1-R_i)/\sum_{i=1}^{j} R_i$
Rank Exponent	$w_i = (n+1-R_i)^z/\sum_{i=1}^{j} R_i^z$

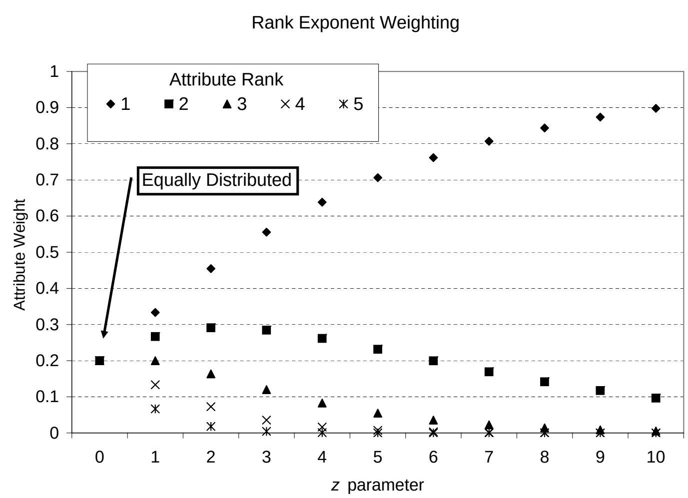


Figure 2.15: Effect of z upon weight

forward with regards to demands placed upon a decision maker and decision analyst. The differences lie within the equations used to ascertain the weights themselves.

Note that Rank Exponent method is simply a variant of the Rank Sum method with $z = 1$). Further, if $z = 0$, then the Rank Exponent methods merely distributes the weights evenly. The parameter z is often estimated from "some convenient pair of attributes (e.g., the most and least important)" [von Winterfeldt and Edwards, 1986, pg. 284]. For the Rank Exponent method, the effect upon the weights, as a result of the parameter z is shown in Figure 2.15.

As an example of the decision rank rule, once the DM provided an estimate of the most important attribute's weight a chart similar to the one in Figure 2.16

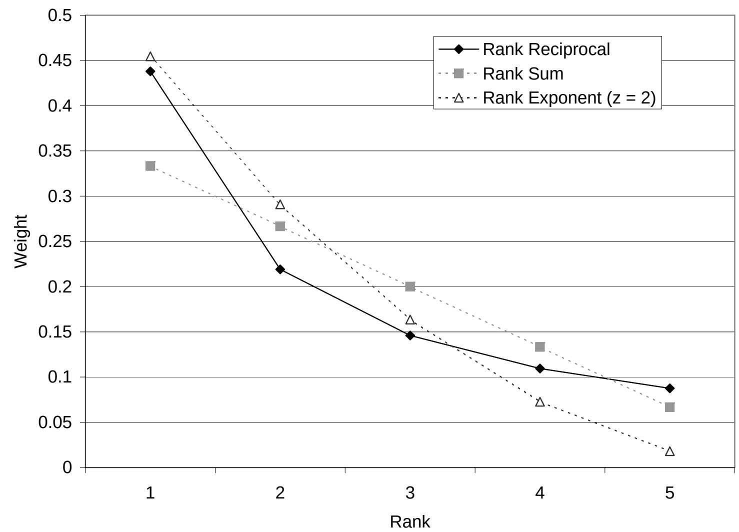


Figure 2.16: Weighting Scheme Comparisons

could be used to compare the ranking schemes, facilitate selection, and illustrate the implications of the scheme to the DM. For example, the rank sum scheme is linear, whereas (comparatively) both the rank reciprocal and rank exponent methods will assign more weight to the higher ranked attribute and possibly less weight to the least ranked attribute. Again, the focus of such approaches should be insight into the problem as a whole.

Although these methods are the easiest to elicit due to their intuitive interpretation, it is important to note that the ranks used as input to these functions (R_i) are ordinal in nature-an inherent limitation in the ranking approach in general. That is "(1) the data classifications are mutually exclusive and exhaustive and (2) data classifications are ranked or ordered according to the particular trait they possess" [Lind et al., 2002, pg. 11]. An example could include ranking targets with priorities (e.g., priority one, priority two, and so forth). Unfortunately, this also implies that even though the attributes may be ranked 1 through j, the calculations might as well operate algebra on 'one,' 'two,' 'three,' ..., and 'j.' In order for these techniques (specifically the mathematical calculations required) to work, the ranking data must

Table 2.15: Scale Types [Narens and Luce, 1986, pg. 168]

Scale	Admissible Transformations	Examples
Absolute	$x \rightarrow x$	John is twice as tall as Bill
Discrete Ratio	$x \rightarrow k^n$, constant $k > 0, n \in \mathbb{Z}$	length in lines of code
Ratio	$x \rightarrow rx, r \in \mathbb{R}^+$	age, speed, Kelvin Temperature
Discrete Interval	$x \rightarrow k^n x + s$, constant $k > 0$, $n \in \mathbb{Z}, s \in \mathbb{R}$	murder rate (based on population proportion)
Log Discrete Interval	$x \rightarrow sx^{kn}$, constant $k > 0$, $n \in \mathbb{Z}, s \in \mathbb{R}$	murder rate : police force (per 1000)
Interval	$x \rightarrow rx + s,\ r \in \mathbb{R}^+,\ s \in \mathbb{R}$	Temperature (Celsius or Fahrenheit), calendar dates
Log Interval	$x \rightarrow sx^r,\ r, s \in \mathbb{R}^+$	density, fuel efficiency in mpg
Ordinal	$x \rightarrow f(x)$, $f$monotonic	beauty, hardness
Nominal	$x \rightarrow f(x)$, $f \in$ 1 - to - 1 functions	names, numbering on athletic uniforms

be at least interval in nature. Lind et al. define interval data as having the same characteristics as ordinal data, plus the "data classifications are scaled according to the amount of the characteristic they possess, and equal difference in the characteristic are represented by equal differences in the measurements" [Lind et al., 2002, pg. 11]. More formal definitions of scale types are shown in Table 2.15.

Therefore, according to Narens and Luce, any monotonic transformation will retain the order of the weights, but does not imply that information regarding the interval between the weights after transformation is meaningful. Despite the lack of a meaningful distance between ranked data, the use of these techniques continues, likely due to the interpretability and repeatability offered to decision makers who may not have extensive training in decision analysis theory [von Winterfeldt and Edwards, 1986, pg. 312]. Interestingly, Barron and Barrett, in their study of several ranking-based techniques (rank ordered centroid, rank sum, and rank reciprocal) concluded that these methods "...represent excellent tradeoffs between ease of assessment and

efficacy of selection of the best or near best alternative" [Barron and Barrett, 1996, pg. 1520].

The other major limitation in these ranking techniques is the explicit involvement of "...the notion of attribute importance [with respect to each other]" [von Winterfeldt and Edwards, 1986, pg. 285]. This, as Keeney and Raiffa suggest, is an inappropriate interpretation of the weights which are (1) solely for the purpose of providing an aggregate value score and (2) directly dependent upon the range of the SDVF [cf. von Winterfeldt and Edwards, 1986, pg. 285].

Unless the decision maker is forced or reminded to take into consideration the attribute ranges during a rank-based elicitation process, the "concept of importance as a basis for weighting" is problematic [cf. von Winterfeldt and Edwards, 1986, pg. 285-6]. von Winterfeldt and Edwards propose the swing-weighting technique instead, and suggest that this approach not only "...counters the criticisms of using extraneous and perhaps even distorting importance judgments..., [but also], given carefully anchored SDVF elicitation techniques is virtually indistinguishable..." from theoretically appropriate indifference methods such as difference measurement, conjoint measurement theory, and weak order models [cf. 1986, pg. 286-7].

In order to implement these techniques in a dynamic environment, application of the rating- and rank-based procedures would likely involve an *a priori* rank ordering specified by the DM (e.g., a ranking of attributes or objectives applicable to the targeting cycle for each phase of the war). The rank reciprocal rule, from Table 2.14, may also be applied in a fashion similar to the approach taken by Pruitt [2003]. For example, in an effort to improve U.S. Homeland Security, the decision model implied a preference for improving capabilities that were at currently inadequate or low levels [Pruitt, 2003]. Subsequently, an alternative that scores high implies that it meets the DM's needs to improve upon current capabilities. For the rank reciprocal rule, these percentages may be substituted for the rankings. A lower percentage level of current capability will result in a higher weight–implying that the DM will

Table 2.16: Example Rating Techniques [Bottomley and Doyle, 2001, pg. 553-554]

Direct Rating	Rates each attribute on a scale of 0 to 100; scores are normalized to produce weights
Point Allocation (Max100)	Assigns the most important attribute a rating of 100; rates subsequent attributes relative to the most important one on a scale of 0 to 99; scores are normalized to produce weights
Point Allocation (Min10)	Assigns the least important attribute a rating of 10; rates subsequent attributes relative to the least important one with no specified scale; scores are normalized to produce weights

prefer to focus first on what objectives need improving the most. As this scheme is dependent upon the current state, it is philosophically but not mathematically similar to the method proposed by Li et al. [2004]. However, a potential limitation of this approach is that, over time, weights would tend to be distributed equally, as opposed to having a distribution representative of their true preferences among the different objectives. At the point where this occurs, the DM should revisit the model and resort to a method that more closely represents his or her preferences.

Three examples of rating methods are described and compared in a case study in Bottomley and Doyle [2001]; the methods are summarized in Table 2.16. Again, in order to implement these in a dynamic weighting fashion, the ratings should be accomplished in advance, with a rating corresponding to a particular phase or period of time during with the decision model will remain applicable.

Another popular family of weight elicitation methods comprises the rating, ranking and swing-weight techniques–SMART, SMARTS, and SMARTER [Edwards, 1977; Edwards and Barron, 1994]. SMART is a ten-step method, described in detail in Edwards [1977] and summarized in Table 2.17. Acknowledging the theoretical ties between attribute preferences, ranges, and the weight values, Edwards and Barron actually recommend against the further use of SMART since "the procedure ignores the fact that range as well as importance must be reflected in any weight" [Ed-

Table 2.17: SMART Methodology [Edwards, 1977, pg. 327-9]

1.	Identify the person or organization whose utilities are to be maximized.
2.	Identify the issue(s) to which the utilities needed are relevant.
3.	Identify the entities to be evaluated.
4.	Identify the relevant dimensions of value for evaluation of the entities.
5.	Rank the dimensions in order of importance.
6.	Rate dimensions in importance, preserving ratios.
7.	Sum the importance weights and divide each by the sum (i.e. normalization).
8.	Measure the location of each entity being evaluated on each dimension.
9.	Calculated the utilities for entities.
10.	Decide.

Table 2.18: SMARTS Methodology [Edwards and Barron, 1994, pg. 307-9]

1.	Identify the purpose and decision makers.
2.	Elicit a value tree.
3.	Identify the entities to be evaluated (alternatives).
4.	Formulate an alternatives-attributes matrix.
5.	Eliminate ordinally dominated options.
6.	Reformulate data from step 4 into single dimensional values (SMARTS assumes all single-dimensional value functions are linear).
7.	Implement swing weighting.
8.	Normalize weights and calculate overall scores.
9.	Decide.

wards and Barron, 1994, pg. 316]. Interestingly, a proposed solution to this initial theoretical shortcoming, SMARTER also ignores this connection. SMART, Point Allocation (Max100), and Point Allocation (Min10) are examples of ratio estimation techniques. To remedy the problems associated with SMART, SMART using Swings (SMARTS) was developed. The steps for this process are summarized in Table 2.18.

Swing-weighting (required for step 7) "...refers to the operation of changing the score of some object of evaluation on some dimension from one value to a different one" in order to compare two hypothetical alternatives [Edwards and Barron, 1994, pg. 316]. The other proposed improvement the authors suggest is SMART Exploiting Ranks (SMARTER) [Edwards and Barron, 1994]. This methodology is identical to

that of SMARTS, with the exception of step 7. In lieu of swing weighting, a rank order centroid approach is taken, which defines the weight set by

$$w_i = \frac{1}{n} \sum_{k=i}^{n} \frac{1}{k}. \tag{2.17}$$

Clearly, other than specific guidance from the analyst to the DM during the elicitation process, the weights derived from SMARTER do not directly account for the associated attribute ranges. Nonetheless, it is easy to see that a multitude of numerical estimation techniques for attribute weights (ranking, direct rating, ratio estimation, and swing weights) offer straightforward estimates of attribute weighting. Although these methods are not without their shortcomings, Stillwell and Edwards investigated and reassessed previous multi-attribute case studies, finding that various rank-based weighting techniques (other than equal weights) as "approximations to ratio weights provided remarkably good results" [Stillwell and Edwards, 1979, pg. 28].

Indifference methods are discussed next, primarily to round out the discussion of weighting methodologies available to the decision analyst. However, these elicitation techniques are more complicated and time consuming. Consequently, their application within a dynamic environment may be even more limited.

2.5.5.2 Indifference Methods

Indifference methods systematically explore either indifference judgments, as seen in [Keeney and Raiffa, 1993], or the strengths of preferences among attributes, illustrated in [von Winterfeldt and Edwards, 1986, pg. 287]. Three classes of these techniques are discussed by von Winterfeldt and Edwards [1986], which are summarized in Table 2.19. As opposed to the numerical estimation methods discussed earlier, the indifference methods yield weights that are ratio in nature. Consequently, further mathematical operations may be justified in the context of dynamic weights.

Table 2.19: Indifference Methods [von Winterfeldt and Edwards, 1986, pg. 287-98]

Method	Description
Difference Measurement	Indistinguishable from SMARTS; assumes independence and additivity of attributes; this approach may be extended to multiplicative models [cf. Kirkwood, 1997, pg. 71-2 for an illustrative example]
Conjoint Measurement	Assumes additive model; does not require strength-of-preference judgments; requires preference and indifference judgments; may be extended to multiplicative models;
Weak Order Model	Requires no underlying assumptions regarding attribute additive or independence; "Exploits the assumption of transitive indifferences to trade multi-attribute alternatives off sequentially until they become comparable;" useful for complex (i.e., non-additive models); extremely difficult in elicitation and expectation of DM to make consistent tradeoffs

Recall that Gabrielli and von Winterfeldt theoretically interpret weights as "how much a (value) unit in one attribute contributes to overall worth relative to a unit in another attribute" [Gabrielli and von Winterfeldt, 1978, pg. 2]. This interpretation confounds the value of the weight with the somewhat arbitrary ranges specified for the single-dimensional value functions [Gabrielli and von Winterfeldt, 1978, pg. 2]. Since the derivation of weights via an indifference method explicitly incorporates information regarding the range of the SDVF, the requisite implicit assumption follows that any change in the range of a SDVF must result in a concomitant change in its associated weight [Gabrielli and von Winterfeldt, 1978, pg. 2]. The following example demonstrates the theoretical connections, and explores areas where the change is not always as expected.

Suppose a value model has two attributes, each of which has a corresponding, linear SDVF function represented as $v_1(x_1) = (x_1)/30, x_1 \in [0, 30]$ and $v_2(x_2) = (x_2 - 20)/40, x_2 \in [20, 60]$. Further suppose that the objectives are preferred such that $v_2 \succ v_1$, which implies that $w_2 \succ w_1$, and the following value tradeoff (taking

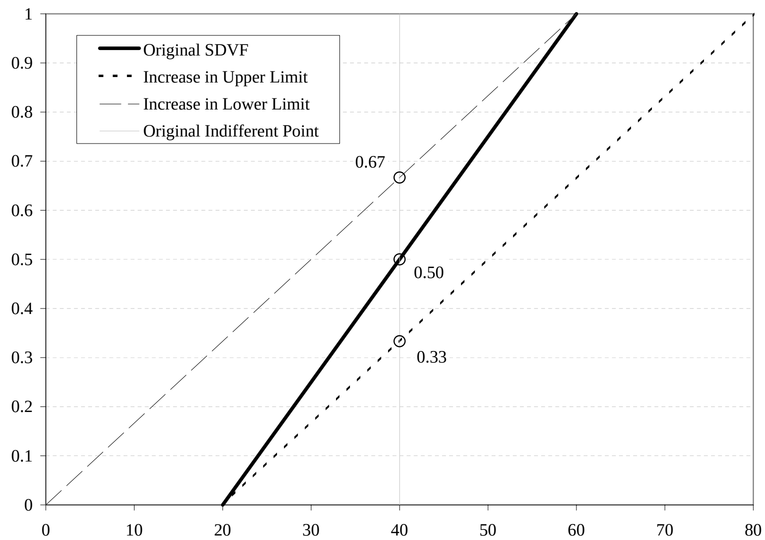


Figure 2.17: Weight Effects due to Range Changes

advantage of requirement 2.12 was made by the decision maker.

$$\begin{aligned} v(0,40) \sim v(30,20) \Rightarrow w_1 v_1(0) + w_2 v_2(40) &= w_1 v_1(30) + w_2 v_2(20) \\ \Rightarrow w_2 v_2(40) &= w_1 v_1(30) \\ \Rightarrow w_2 &= 0.5 w_1 \\ \Rightarrow \mathbf{w} &= \left[\frac{1}{3}, \frac{2}{3}\right] \end{aligned}$$

In order for the expected phenomena to occur, the SDVFs involved must (1) both be anchored at a natural or proxy zero that cannot decrease and (2) are both either monotonically increasing or monotonically decreasing. Assuming that the value of x_2' remains the same regardless of the range change, several cases of the SDVF for $v_2(x_2)$ are exhibited in the Figure 2.17.

As shown in Figure 2.17, an increase to the right in the range increases the value of w_2, shown by

$$w_2 v_2(40) = w_1 v_1(30) \Rightarrow 0.33 w_2 = w_1 \Rightarrow \mathbf{w} = \left[\frac{1}{4}, \frac{3}{4}\right].$$

However, an increase to the left in the range decreases the value of w_2, given by

$$w_2 v_2(40) = w_1 v_1(30) \Rightarrow 0.67 w_2 = w_1 \Rightarrow \mathbf{w} = \left[\frac{4}{10}, \frac{6}{10}\right].$$

Experimentally, von Nitzsch and Weber note that a number of "empirical studies indicate that weights are not adjusted properly to changes in range" due to the decision maker, the analyst, the elicitation process itself, or a combination of these elements [von Nitzsch and Weber, 1993, pg. 937-8]. The authors attempt to capture to what degree this type of error promulgates within the decision analysis process. Once a value function is elicited using a given scale (worst to best or endpoints), changing the scale is often problematic, which leads to the tendency to "choose end points very likely to include any possible future alternatives" [von Winterfeldt and Edwards, 1986, pg. 230]. von Winterfeldt and Edwards also recommend the use of an 'acceptable' range, described by the relationship: *actual* $\subseteq$ *acceptable* $\subseteq$ *available* $\subseteq$ *theoretically feasible* [von Winterfeldt and Edwards, 1986, pg. 230-1].

The model of interest within the experiment conducted by von Nitzsch and Weber was an additive value model under certainty, comprised of linear value functions. They studied three ranges: an initial range based upon the decision makers intuition, a smaller range half of the initial range, and a larger range twice the initial range. For each range, the authors elicited weights via direct-ratio, described in [Edwards, 1977, pg. 328], and a regression technique called conjoint analysis that explores implicit tradeoffs between attributes during the decision maker's evaluation of hypothetical alternatives [von Nitzsch and Weber, 1993].

The authors mentioned the value trade-off recommended by Keeney and Raiffa [1993], but noted that "the derivation of weights from these statements can be done to guarantee the range sensitivity to be equal to one" [von Nitzsch and Weber, 1993, pg. 939]. This implies that the value trade-off approach is the most theoretically appealing when trying to ensure that the decision maker accurately accounts for

changes in ranges during the development and communication of the weights. They mathematically define range sensitivity as $s = (m^{emp} - 1)/(m - 1)$.

This measures percentage change in range accounted for by the decision maker during the development of the new weights ([von Nitzsch and Weber, 1993, pg. 938-9]. Using an increase in the range of attribute 1 as an example, m is defined by the ratio in Equation 2.18. [von Nitzsch and Weber, 1993, pg. 938].

$$m = \left(\frac{w'_1}{\sum_{j=1}^{n} w'_j}\right) \Big/ \left(\frac{w_1}{\sum_{j=1}^{n} w_j}\right) \tag{2.18}$$

The modified weights (w') result from using essentially the original information elicited to construct the initial weights (w). For example, if preference trade-offs were used, the original indifference point elicited would be used to calculate the new weights associated with the larger range for attribute 1. Once the ranges are communicated to the decision maker, weight elicitation is re-accomplished, new weights (w'') are calculated, and the final piece of information is available for Equation 2.19 [von Nitzsch and Weber, 1993, pg. 938].

$$m^{emp} = \left(\frac{w''_1}{\sum_{j=1}^{n} w''_j}\right) \Big/ \left(\frac{w_1}{\sum_{j=1}^{n} w_j}\right) \tag{2.19}$$

Note that m and m^{emp} are essentially the theoretically required and empirically observed changes in weights respectively [von Nitzsch and Weber, 1993, pg. 938]. Ultimately, von Nitzsch and Weber concluded that both weighting methods employed resulted in $s < 1$, implying that "subjects only partially adjusted their weight judgment to the change in range," and that this "process was especially bad" when the direct ratio elicitation method was used, often resulting in biased weights [von Nitzsch and Weber, 1993, pg. 942].

In practice, Lootsma suggests that "criteria have emotional or social values which neither depend on the actual decision problem itself nor on the method of

analysis" [Lootsma, 1999, pg. 33]. Therefore, it appears to simply be a human tendency to infer a certain level of relative importance from the values of the weights. In their early experimentation, using an approach similar to that adopted by von Nitzsch and Weber [1993], Gabrielli and von Winterfeldt note,

> ...that people can give importance orderings without specified alternative and ranges may mean that they have some plausible set of alternatives and ranges in mind, when judging importance. According to this interpretation the importance judgments should only change when the environment radically changes the plausible set of alternatives [1978, pg. 28].

This finding is also observed by Bottomley and Doyle, who state that "intuitive weights reflect a subject's general attitude towards an attribute, and an implicit range of outcome values, ..." thus enabling decision makers to specify attribute preference (weights) without specific knowledge of attribute ranges [Bottomley and Doyle, 2001, pg. 554].

2.5.6 Weighting Issues Summarized

With respect to the requisite change in weight as a result in the change of attribute range, the gap between the theoretical and practical results remains. Keeney cites "assessing value trade-offs independent of the range of consequences" as one of the 'top 12' mistakes in decision analysis [Keeney, 2002, pg. 940]. The results of Gabrielli and von Winterfeldt indicate that subjects found it difficult to adhere to the theoretical requirements in practice and attribute these findings to either an inherent flaw in their test problem or the complexity of the task (or both) [1978, pg. 20, 22]. Unfortunately, in an effort to resolve this issue, another experiment led them to conclude that... "Even in an absurdly simple problem subjects apparently had problems appreciating the sensitivity of importance weights to a change in the range of an attribute" [Gabrielli and von Winterfeldt, 1978, pg. 25-7].

Pöyhönen and Hämäläinen, who provide a more recent study of a variety of weighting techniques, not only arrive at this same conclusion, but also suggest that weight values are more dependent upon the number of attributes rather than the elicitation process [2001, pg. 581]. Resolving this long-standing issue is beyond the scope of this research. Ultimately, any modeling effort should be undertaken in order to improve the fundamental understating of the decision problem, in which the weighting process provides the most benefit [Hämäläinen and Salo, 1997, pg. 340]. As there are several interpretations of weights, there are several techniques used to elicit them. The general context in which these weighting techniques will be applied is presented next, followed by example applications of various dynamic weighting schemes.

2.5.7 The Dynamic Decision Environment

> Space is more or less tangible and/or visible, but time and preference are volatile. Living creatures have a surprising ability, however, to control a time-dependent series of rhythmic actions like walking, running, and tapping, which are controlled by a timekeeper. Many living creatures also have a biological or physiological clock to measure the time which elapsed since a particular moment. So, if time can subjectively be measured, the gradations of preference may be measurable as well [Lootsma, 1999, pg. 9].

As Lootsma suggests, the theoretical underpinnings of the field of decision science not only deals with preference, but time-dependency of preference as well. Indeed, he argues that the values and weights are situation-specific, essentially becoming irrelevant once the decision of interest is made [Lootsma, 1999, pg. 33]. He also presents the counter argument in that decision makers seek consistency "...over a coherent collection of decision problems" [Lootsma, 1999, pg. 33]. Dynamic weights, then, appear to lie within these bounds, their values dependent upon time, the decision environment (and consequently the collection of decision problems addressed), or both [cf. Keeney and von Winterfeldt, 1989, pg. 86]. Either way, the common denominator for both perspectives is the assumption that the weights are

valid for a given period of time, either instantaneously or during a phase of closely related decision problems.

Edwards posits two general types of decision theory–static and dynamic [1962, pg. 59]. *Static* represents the traditional, one-time use of a decision model garnered from approaches such as multi-objective decision analysis, value focused thinking, multi-attribute utility analysis, and others. *Dynamic* represents the enactment of a sequence of related decisions, where each subsequent decision may (or may not) benefit from either the results or the information obtained via the consequences of the previous decision [Busemeyer, 2002, pg. 3903], [Edwards, 1962, pg. 59-60]. Interestingly, Busemeyer notes categories of decisions within this realm include "...fighting fires, navigational control, battlefield decisions, medical emergencies, etc." [Busemeyer, 2002, pg. 3903]. Busemeyer also suggests that dynamic decision making is characterized by three features:

1. A series of actions must be taken over time to achieve some overall goal;
2. The actions are interdependent so that later decisions depend on earlier actions; and,
3. The environment changes both spontaneously and as a consequence of earlier actions [Busemeyer, 2002, pg. 3903].

The notion of dynamic decision making plays an important role in Bayesian networks, decision support systems, and expert systems. Mussi discusses the incorporation of utility theory (value models under uncertainty) to measure (and compare) consequences analyzed in such systems and includes an approach to dynamically weight the models used to facilitate selection of the next course of action [Mussi, 2004, pg. 95-6]. However, the dynamic nature of his model is essentially the selection of a pre-determined weight set (elicited in advance), based upon the current system state [Mussi, 2004, pg. 95]. Figure 2.18, adapted from Busemeyer [2002], illustrates the general flow of information, feedback, and uncertainty within the dy-

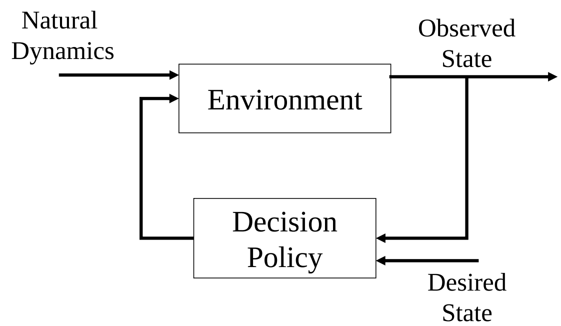


Figure 2.18: Dynamic Decision Environment [Busemeyer, 2002]

namic decision environment. The environment is (potentially) affected by general uncertainty as well as decisions made in the past-a complicating factor not found in the static realm of decision theory [Edwards, 1962, pg. 60]. Ultimately, a decision maker will implement a decision policy, enacting decisions to shape the environment, in order to 'close the gap' between the current, observed state and a desired end state.

Another study somewhat related to this research is the work of Weisbrod et al. [1977]. Weisbrod et al. developed a decision aid that complemented a decision maker's decision process by incorporating information from a simulated environment. The next set of calculations was predicated upon the previous actions which determined what information would be available at the time of the next decision. However, this approach focused upon an expected utility or probabilistic approach rather than the deterministic assumption taken here Weisbrod et al. [1977].

A more closely related approach is described by Li et al. [2004], who interpret the weights as indicators of relative importance, develop the transformation from a single weight vector to one that is dependent upon the current system state Li et al. [2004, pg. 163-5]. Despite the extremely detailed development of the fuzzy math underlying the "state dependent weight vectors" they propose, neither a means to elicit nor a means to tie these states directly to the environment is suggested Li

et al. [2004, pg. 168-78]. In his research in dynamic decision making, Busemeyer concludes the following:

> [Subjects] who perform best are those that set integrative goals, collect systematic information, and evaluate progress toward these goals. Subjects, who tend to shift from one specific goal to another, or focus exclusively on only one specific goal, perform more poorly [Busemeyer, 2002, pg. 3907].

Therefore, the overarching assumption is that a decision model is developed that captures a decision maker's preferences in extraordinary detail, but at extraordinary cost. The multi-objective decision modeling approach strives to avoid the latter reason for poor performance while careful selection of the time period to which this model applies will help avoid the former reason for poor performance. As opposed to a 'one-time' decision, suppose also that this model could be used to evaluate alternatives in the same decision context, but in a different time period, allocating potentially different weights to the objectives. In order to clarify the decision problems and contexts in which this approach may apply, a few illustrative examples are provided.

2.5.8 Examples of Interest

Although value models are not explicitly constructed, the following examples are intended to illustrate the decision situations that may yield deterministic multi-attribute value models applicable to a series of related decision problems, with the only change in structure being the weights over time.

2.5.8.1 Company Valuation

A recent improvement was proposed to the process of fundamental analysis, which examines "the underlying forces that affect the well being of the economy, industry groups, and companies...," with the goal of predicting future performance

Table 2.20: Levels and Focus of Fundamental Analyses

Level	Areas Examined
Company	Financial data, management, business concepts, and competition
Industry	Forces of supply and demand
National	Economic data permitting assessment of present and future economic growth

and profitability [Anonymous, 2005]. This analysis typically involves three economic components, summarized in Table 2.20.

DeGraw noted that these types of analysis are generally based upon the two following assumptions–"that each of the three components is weighted equally and that their relative importance doesn't change over time" [DeGraw, 2001, pg. 78]. However, he notes that "Since [initial public offerings] are smaller, less liquid and considerably more dynamic, industry differences appear to be considerable, and the relative influence of the valuation components appears to vary across industries" [DeGraw, 2001, pg. 78]. This is an ideal example of the underlying situation assumed for the proposed dynamic weighting approach. The result is a decision model with objectives applicable to the range of alternatives (initial public offerings to blue-chip companies) that can accommodate different weights based upon the context of the analysis (i.e., the time-line associated with the company's level of financial maturity). Of course, it could be argued that a different model could be developed for each industry's level of maturity. However, in defining such models, their specificity will naturally limit their applications as well incur more development and analysis costs in the process. Another potential modeling situation, also financially oriented, is consumer preference.

2.5.8.2 Preferences for Consumption

Consumer demand for various goods and services may also fit the genre of dynamic weighting methodologies. For example, the distributions associated with

portfolio asset allocation will (generally) change over time based upon the preferences for a given objective (capital gains, steady growth, etc.) and its corresponding value in the context of that investor's age. Similarly, building upon the classic 'choose the best automobile' example, there are clear areas where time will affect the distribution of weights, given that the remaining structure of the model remains constant. For example, consider the objectives "comfort and refinement, fuel consumption, safety and security features, ride and road handling, performance, aesthetic appeal, reliability, running and maintenance costs, and space and practicality" examined by [Bottomley and Doyle, 2001, pg. 555].

An excited, newly licensed driver (also likely to be a hormonal teenager) is prone to value performance-related objectives over comfort and practicality. Whereas, the older, more mature driver, particularly one with a family, may value safety- and practicality-related objectives over those associated with high-performance vehicles. (An exception to this may involve an individual enduring a 'mid-life crisis,' during which their values would revert back to that of the teenage decision maker.) Next, a military example is posed, where the time-lines potentially associated with a dynamic weighting approach occurs on much shorter intervals-minutes to months, as opposed to years.

2.5.8.3 Joint Targeting Cycle

According to Joint Publication (JP) 3-0, the joint targeting cycle (JTC) has six phases:

1. Commander's Objectives, Guidance, and Intent;
2. Target Development, Validation, Nomination, and Prioritization;
3. Capabilities Analysis;
4. Commander's Decision and Force Assignment;
5. Mission Planning and Force Execution; and,

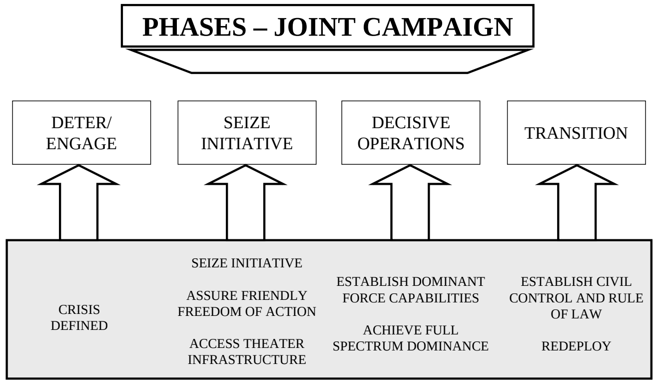


Figure 2.19: Phases of the Joint Campaign [DOD, 2001, pg. III-19]

6. Combat Assessment [DOD, 2001, pg. III-28].

Although a decision model that takes advantage of dynamic weighting may have application in all phases of the JTC, phases (4) and (6) are specifically discussed.

Within the "Commander's decision and force assignment" phase, available assets are assigned to various missions based upon operational, and generally time-dependent, needs. This assignment is accomplished through a process called *apportionment*, defined by Air Force Doctrine Document (AFDD) 2-1, *Air Warfare*, as follows.

> Apportionment is the determination and assignment of the total expected aerospace effort by percentage, priority, weight of effort, or some other appropriate means, that should be devoted to the various aerospace operations and geographic operations for a given period of time [USAF, 2000, pg. 50].

Generally, the time periods are either the duration (or a subset thereof) of a phase in the campaign. These phases are highlighted in Figure 2.19. The process of air apportionment assists Joint Force Commanders in ensuring "...the weight of the joint force air effort is consistent with campaign phases and objectives" [DOD, 2001, pg. III-29].

The efforts to which these weights are assigned (apportioned) are defined as functions, which comprise the "broad, fundamental, and continuing activities of

aerospace power" (e.g., Counter-air, counter-space, counter-sea, counter-land, strategic attack, and counter-information, among others) [USAF, 2000, pg. 5]. As progress is made throughout the phases of the campaign, there will be requisite shifts in levels of effort placed toward these functions. Ultimately, major changes in weights (i.e., apportionment) may occur at each campaign phase. However, due to the dynamic nature of war and the interaction with an adversary that has goals and objectives contrary to our own, the associated weights may change even more frequently. The feedback process of the phase (6) of the targeting cycle, combat assessment, will likely employ a dynamic weighting process that is dependent upon the operations tempo, which involves time periods from minutes to weeks.

Combat assessment (CA) is the evolution of the traditional process of battle damage assessment (BDA) within today's complex battlefield, and is formally defined as "the determination of the overall effectiveness of force employment during military operations" [DOD, 2001, pg. IV-17] The underlying uncertainty involved in this type of information lends itself to multi-attribute utility models instead of their deterministic counterpart–multi-attribute value models. Nonetheless, CA plays a vital role in the use of limited resources (e.g., fuel, sorties, time to complete objectives, weapons, and so forth) and essentially affects the apportionment process to ensure that the commander's objectives are met. One possible use would be to improve the Time Critical Targeting (TCT) process. TCT occurs when a Time Sensitive Target (TST) (e.g., location of enemy leader becomes known) is found and the ATO needs to be updated to put resources on the TST.

2.5.9 Conclusions Regarding Weighting

Indifference methods offer a theoretically sound approach to establishing attribute weights that not only account for attribute importance, but also accommodate the range of the attribute's SDVF as well. However, this section has highlighted several studies that suggest decision makers have difficulty in responding with appro-

priate weight changes as a result of changes in ranges. Additionally, several studies have determined that the overall decision recommendation is somewhat insensitive to the preference indifference methods recommended by Keeney and Raiffa [1993] and other, less theoretically-pure methods often categorized as numerical estimation techniques [cf. Pöyhönen and Hämäläinen, 2001].

The ease of use suggests that numerical estimation techniques are likely more suited to dynamic weighting in an operational setting. Further, several researchers have concluded that "None of the more complicated weighting procedures performed any better than the simple technique of directly assessing the rank ordering and arithmetically transforming the ranks into weights" [John et al., 1980, pg. 22], [cf. Stillwell and Edwards, 1979, pg. 28]. Overall, "The maxima of utility theory are very flat, which means that modest errors in changing numbers are unlikely to affect orderings" [Gabrielli and von Winterfeldt, 1978, pg. 20]. As always, it is important to remember that the purpose of decision analysis is to provide insight. Through the process of making judgments explicit, "it is easier to identify weaknesses in the reasoning behind a decision" [Keeney and von Winterfeldt, 1989, pg. 86]. An automated (or semi-automated) method to dynamically weight value models may limit the benefits associated with the process-critical thinking. Then again, thinking about a particular decision problem within current and future contexts may prove beneficial to the analysis process, yielding an advantage to those who have a dynamic process implemented within today's dynamic environment. Ultimately, higher level trade-offs must be made between the DM's time, their amount of involvement, and the acceptance of the assumptions regarding the nature of weights over time.

Dynamic weighting for several weighting methodologies is developed in Chapter VI; such techniques would lend themselves to the dynamic scenarios discussed earlier. This concept is further leveraged for the aggregation, as required, of multiple, contextual layers of social networks in order to facilitate other computational analyses. The next section discusses mathematical programming formulations that

may serve as a means to quantitatively analyze the networks *in situ* as well as the effects of influence operations applied to the network.

2.6 Mathematical Programming Approaches

Mathematical programming (MP) plays an important role within this research, which extends some of the concepts initially developed by Renfro [2001] and applies this modeling technique to other, traditionally sociological types of problems. These efforts further the analysis capabilities within this field of research. For example, slight modifications of the adjacency matrix serve as a direct input into a variety of mathematical programs, thereby offering prescriptive analysis capabilities. As discussed earlier, a modified version of the node-edge incidence matrix of the same social network is useful in studying the literal flow of influence [Clark, 2005; Renfro, 2001] as well as estimation for the potential of influence flow to ascertain actor centrality Freeman et al. [1991]. Extensions, both theoretical and applied, of these works are included within the primary research goals. Several mathematical programming formulations and their roles within this research are briefly reviewed. Other than the works previously mentioned, very limited substantive connections between mathematical programming and social network analysis have been made.

Recall that a given social network is typically described by a graph, $G = (N, E)$, where N is the set of nodes (or individuals in this setting) within the network of interest and E is the set of relations upon which the context of G has been constructed. For example, if G is the network of 'who knows whom,' then an edge or relation, $e_{ij} \in E$ implies that individual i knows individual j [Wasserman and Faust, 1994, pg. 150]. The sociomatrix, denoted $\mathbf{X}$, is one of the primary tools used by sociologists and is equivalent to the network's corresponding adjacency matrix. If actor 1 is adjacent to actor 2 within a particular context of study, then $x_{12} = 1$, zero otherwise [Wasserman and Faust, 1994, pg. 150]. Generally, $\mathbf{X}$ is symmetric, but asymmetric relationships can be indicated whenever $x_{ij} \neq x_{ji}$.

In addition to direction, sociometric data may be valued, rather than simply indicating the existence (1) or non-existence (0) of a contextual relationship between two individuals within a social network. This necessitates a valued, directed or undirected, graph to accurately represent the social network, as opposed to the symmetric and dichotomous relationship oftentimes assumed in past sociometric studies. As Buchanan noted, "In a social network, the bonds between good friends are not the same as those between weak acquaintances" [Buchanan, 2002, pg. 145]. Means to numerically estimate the potential of such bonds has been posited by Renfro [2001] via a social closeness function–the stronger the bond, the greater the value of social closeness. More recently, Clark [2005] suggested another potential influence function based upon multi-network structures as well as external, demographic-oriented data of the individuals comprising the network.

This suggests that potential opportunities lie within the theoretical improvements to be made in the merging of these two sciences, sociology and operations research. Each problem of interest is summarized and followed by findings available in literature, if any, and potential applications of these techniques in the endeavor of studying social networks are posited.

2.6.1 Minimum Spanning Tree Problem

Since several elements of this research deal with network flow formulations, spanning trees in general are clearly of interest due to the relationship between rooted spanning trees and non-singular bases in network flow programs [Ahuja et al., 1993, pg. 450] [Nemhauser and Wolsey, 1999, pg. 77]. Renfro also posited the use of spanning trees, from either a minimum or maximum perspective, to ascertain the social connectivity of members within a network, using social closeness as the arc weights [Renfro, 2001, pg. 48].

As discussed earlier, Borgatti defines KPP-2 as the subset of members maximally connected to the entire network [Borgatti, 2003a, pg. 2]. The impetus behind

Borgatti's key player problem is realization that some traditional SNA measures, such as closeness or betweenness, that attempt to measure the importance of a specific individual do not translate well when a subset of the individuals are of interest [cf. Wasserman and Faust, 1994]. Consequently, improving upon this methodology by combining the measures developed by Renfro and Clark, as well as those within this research, may allow solutions to improved, 'operational' constructs of the key player problem. The minimum spanning tree and forest concepts are extended to KPP-2 to achieve this goal.

Note also that the KPP-2 concept may be abstracted beyond that of social networks. Suppose, for example, that the branches of the maximal spanning social tree served as communications- or influence-interdiction targets. Diffusion of innovations, rumor theory, and other related literature and methodologies may benefit from the use of such a tree as an initial starting point from which to generate an external influence upon a network. For example, if the objective were to reach as many individuals within a network as possible, targeting the well connected individuals or frequently used (or heavily relied upon) communications channels would be of interest. However, individuals could comprise populations and networks of communications channels could be comprised of major cities, street intersections within a town, popular web pages, and so forth.

2.6.2 Covering and Partitioning Problem

Borrowing from Nemhauser and Wolsey [1999, pg. 6-7], a problem with a constraint set $\mathbf{Ax} \geq \mathbf{1}, \mathbf{x} \in \{0,1\}$ ($\mathbf{Ax} = \mathbf{1}, \mathbf{x} \in \{0,1\}$) is generally referred to as a covering (partitioning) problem, respectively. It will be shown that a modified version of the reachability matrix results in a covering problem that solves the KPP-2 problem. This MP approach provides several benefits over the heuristic approach developed by Borgatti. These benefits include: a guaranteed optimal solution, incorporation of directed networks, ability to incorporate valued relations, ability to

Table 2.21: GFP Variable Definition

Variable		Definition
c_{ij}	$\equiv$	the cost per unit flow induced from node i to node j
x_{ij}	$\equiv$	number of units of flow from node i to node j on arc (i, j), $x_{ij} \in [0, u_{ij}]$
b_i	$\equiv$	0 if node i is a transshipment, or 'pass-through,' node; < 0 if demand is required by node i; and, > 0 if supply is provided from node i
g_{ij}	$\equiv$	a rational value $> (<) 1$ that indicates if arc (i, j) is gainy (lossy); if $g_{ij} = 1$, then the arc (i, j) is neither one
N	$\equiv$	the set of nodes (individuals) within the network
A	$\equiv$	the set of arcs (i, j) (connections between individuals) that form the network

discount actors not reachable by external influences, ability to encompass multiple dimensions of relationships, and so forth. Further modifications, leveraging the use of slack variables, may yield a partitioning problem that permits the selection of a set that attains a percent goal, such as a subset, of the population reached. This mirrors an aspect of Borgatti's research and offers another element of flexibility to the MP approach.

2.6.3 Generalized Network Flow

As seen in [Renfro, 2001; Clark, 2005; Freeman et al., 1991], network flow models provide a useful methodology for the study of influential actors within a network. These works, however, focus upon the development of interpersonal measures that serve as capacities of inter-personal influence. Suppose in addition to this information, inter-personal measurements provided differential influence assessments–specifically, gains, losses and thresholds of persuasion. The generalized network flow problem (GFP) has the ability to mathematically accommodate such phenomena. Variants of this problem include the maximum flow and minimum cost formulations. Ahuja et al. [1993] provide a general formulation for the GFP. Variable definition and problem formulation are briefly reviewed below [Ahuja et al., 1993, pg. 567-8].

$$Minimize \quad \sum_{(i,j)\in A} c_{ij}x_{ij} \tag{2.20}$$

$$\sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(j,i)\in A\}} g_{ij}x_{ij} \geqslant b_i \ \forall i \in N \tag{2.21}$$

$$0 \leq x_{ij} \leq u_{ij} \ \forall (i,j) \in A \tag{2.22}$$

The objective function, given by Equation 2.20, seeks to minimize the total cost of flow through the network, subject to the constraints, given by Equations 2.21 and 2.22. Constraint 2.21 is an extension of the *mass balance constraint* that allows for potential violations of traditional conservation of flow assumptions. This extension, a relaxation of the original formulation of Equation 2.21, replaces the equality, "=", with "≥" as shown. Typical minimum cost maximum flow problems further specify that the sum of flow entering the system must equal the sum of flow exiting the system. This relaxation facilitates feasibility, particularly when gains and losses affect flow (e.g., 1 unit enters and, due to gains, say 2 or more must exit) and when arcs are capacitated (e.g., there exists a maximum amount of flow that may traverse the arc-the social closeness serving as an upper bound in this case).

As previously mentioned, a variant of the GFP is the maximum flow problem. The formulation, modified from [Ahuja et al., 1993, pg. 168] to incorporate gains and losses, follows.

v ≡ value of the flow (the sum of all sources must equal the sum of all sinks)

s ≡ value of the flow provided by a source

t ≡ value of the flow demanded by a sink

$$Maximize \quad v \tag{2.23}$$

$$\sum_{\{j:(i,j)\in A\}} x_{ij} - \sum_{\{j:(j,i)\in A\}} g_{ij}x_{ij} = \begin{cases} v & \forall i \in s \\ 0 & \forall i \in N - \{s \cap t\} \\ -v & \forall i \in t \end{cases} \quad (2.24)$$

$$0 \le x_{ij} \le u_{ij} \;\forall\, (i,j) \in A \quad (2.25)$$

This formulation is useful in determining whether or not a flow between two targets (or target sets) of interest through the network is even possible, implying that the course of action (to induce a flow of influence between two or more nodes of interest) offers an opportunity for achieving its objectives.

The GFP and its variations provide a means to deal with a variety of real-world problems. Various processes that undergo degradation or improvement over time or distance may be modeled. As an example, picture a ditch irrigation system. In hot weather, evaporative processes diminish the volumes of water as it flows though the system. Alternatively, precipitation (rain, snow, etc.) may increase the volume of water. These are losses and gains in flow, respectively. Further, ignoring losses and gains for a moment, at any point in the system, a junction, the law of conservation of mass dictates that the amount of water flowing into the junction must equal the amount flowing out. Gains and losses at a junction are then modeled by the right hand side, where ($b_i > 0$) and ($b_i < 0$) respectively. Using the work of Renfro [2001] and Renfro and Deckro [2003] as a launching point, continued development of parallels between the flow of commodities in the physical world and the flow of influence in the behavioral realm are sought.

A recent work delineating the types of flows through social networks offers a complicating factor to several network formulations. Borgatti developed a typology of flow processes observed in social networks for various 'commodities.' Gossip, for example, "spreads by replication rather than transference . . . [and] normally does not pass the same link twice, but can pass the same node multiple times." [Borgatti, 2005, pg. 57] Current formulations as described above of network flow within social

networks are incapable of modeling this phenomenon and others within Borgatti's typology. This is primarily due to the mass balance constraints as well as the nature of the solutions in general.

2.6.4 P-Median Problem

Another potential approach to apply to the KPP-2 issue is that of the p-median problem. Suppose key players are viewed as facilities that serve, or influence, themselves and other members not within the key player set–all of which are viewed as customers. The objective, in general, is to minimize demand-weighted distance between the facilities and customers. Let d_{ij} be the distance from actor i to actor j; let $(X_j = 1)$ if actor j is selected as a key player, zero otherwise; let $(Y_{ij} = 1)$ if actor i is 'influenced' by actor j, zero otherwise; and, let P be the size of the kp-set. Note that, due to the nature of the variables, both symmetric and asymmetric graphs can be evaluated. Additionally, distances no longer need be limited to two steps away. In fact, any non-zero distance may be incorporated, as well as weighted (e.g., with social closeness or aggregated multiplex values) networks are available for analysis. Assuming the demand is a constant (unity), the formulation is as follows. (Modified from [Bozkaya et al., 2002, pg. 180-1])

$$\text{(P-Med)} \quad \texttt{Min} \sum_i \sum_j d_{ij} Y_{ij} \tag{2.26}$$

$$\texttt{Subject To} \quad \sum_j Y_{ij} = 1 \ \forall \ i \tag{2.27}$$

$$\sum_j X_j = P \tag{2.28}$$

$$Y_{ij} - X_j \leq 0 \ \forall \ i, j \tag{2.29}$$

$$Y_{ij} \in \{0,1\} \forall \ i, j; \quad X_j \in \{0,1\} \forall \ j \tag{2.30}$$

Constraint 2.27 ensures that each actor is assigned to only one key player; Constraint 2.28 specifies the number of key players to be used; and Constraint 2.29 ensures that

an assignment cannot be made to an individual j unless that individual is indeed a key player. Various extensions of this formulation could be applied. Unfortunately, unless advanced techniques are applied (e.g., Lagrangian relaxation), solving a large problem of this nature to optimality may be computationally challenging.

2.6.5 Disconnecting Sets

Recalling the KPP-1 definition and the current interest in disrupting various aspects of terrorist networks–such as the financial, support, and communication layers–disrupting the overall functionality of a non-cooperative network by disconnecting it into distinct components (e.g., isolating cells from one another to impede planning efforts). Several works have examined disconnecting sets in the context of adversarial networks, physical and social. The research of Jarvis [1968] and Greenberg [1968] discuss aspects of optimal attack strategies of command and control networks, both from a multi-commodity perspective. More recently, Leinart [1998]; Leinart et al. [2002] and Pinkstaff [2001] have explored essentially the same topic, but from a cut-set enumeration and valuation approach, generally focusing on a single-layered network. Lastly, Kennedy developed a methodology to evaluate cut sets and optimize the targeting process for multi-layered infrastructure networks. [Kennedy, 2003]

Renfro posited two forms of multi-contextual analyses: multi-criteria and multi-commodity. The former incorporating relational aspects or contexts independently, whereas the latter considers multiple contexts but assumes that all contexts share a given capacity of potential influence between any two individuals. [Renfro, 2001, pg. 67] Such concepts are of interest due to the layered network approach of this research.

Unfortunately, the transition from single- to multi-commodity networks poses some challenges, as the max-flow, min cost cut set easily found for the former does not generalize to the latter. [Jarvis, 1968, pg. 40] In addition, "...in general, for

multi-commodity graphs the minimum disconnecting set is not necessarily a cut-set" [Jarvis, 1968, pg. 40]. Given a graph with m nodes and r commodities, enumeration of all chains of all possible lengths can equal $r2^{(m-n+1)}$, resulting in cut-set verification or enumerative methodologies computationally problematic when applied to large single-commodity or reasonably-sized multi-commodity networks [Jarvis, 1968, pg. 47]. Further, the technique of replacing an undirected arc with an equivalent set of two directed arcs does not extend well to the multi-commodity network problem as "...only the forward arcs contribute to the capacity of a single-commodity cut-set" [Greenberg, 1968, pg. 13-15].

2.6.6 Path Enumeration Techniques

Although path (or chain for undirected graphs) enumeration may be computationally expensive, there may be reasons that support such an endeavor. For example, the promising approach by Stephenson and Zelen [1989], described earlier, is capable of analyzing valued graphs, but is limited to symmetric networks. This essentially eliminates the possibility of incorporating gains and/or losses, as well as aggregate values capturing multiple dimensions that are not necessarily symmetric in nature. The method is unique in that it avoids path enumeration. However, it is posited that since explicit path enumeration of a symmetric, dichotomous graph is demonstrated to be equivalent to the information centrality measure, then explicit path enumeration may provide a more flexible (but potentially computationally intractable) measure. Several path enumeration methods exist-from one node to all others Misra and Misra [1980], all paths of a given length Parthasarathy [1964], and all paths between two specified nodes Migliore et al. [1990], are just a few prominent examples within the literature.

Clearly, given the current context, operations research offers an array of models that can be used to analyze social networks.

2.7 Summary

To this point, a number of methodologies and theories dealing with both open and non-cooperative social networks have been presented. These works include a variety of Operations Research, Sociological, and Behavioral Theory efforts, all of which provide the bases for this research. The overall goal of which is that new and useful theory, and concomitant methodologies, describing and analyzing social networks of non-cooperative organizations will be realized. Given the improved understanding and insights provided by the proposed research, decision makers can then be offered better courses of action that impute influence upon the network in order to achieve a target influence, perception, or outcome to one or more actors within the network through either direct or indirect means. The next chapter presents an overview of the methodology proposed to accomplish these tasks.

III. Research Approach

3.1 Overview

This chapter outlines the research activities seeking to improve the theoretical and methodological approaches available to analyze layered social networks. A series of approaches are developed to investigate aspects of social networks. The layered concept primarily provides a means to (1) derive a measure of relationship strength and (2) offer insight into potential courses of action that may increase the fragility of the target network or disrupt it entirely through the use of information operations. In addition, the use of exogenous data characterizing the individuals within the network to ascertain power or persuasion differentials is explored. This measure of persuasion, also referred to as influence, of one individual over another is used as a multiplier in generalized network flow applications, again to explore various courses of action as well as supporting a new social network analysis measure.

Recall the layered network diagram presented in Chapter I and repeated in Figure 3.1. In this figure, each layer represents a context within which the actors may or may not be affiliated. Context examples could include familial relationships, training camps attended together, known friendships, business interactions, known animosities, resources shared, specialized skills or training, and so forth. For each layer or context, if an actor is connected to any other actor in the same context, those actors and that link are recorded. Given that the network of interest is comprised of N individuals, each layer can include no more than the same set of N individuals. Examining the diagram from directly overhead, an actor appearing in more than one context would be aligned vertically, as noted by the 'unique actor' indicator.

It is assumed that human interaction simultaneously accounts for multiple, underlying relationships or contexts within which those relationships were developed. For example, two complete strangers may be treated differently based upon the known contexts that comprise a newly formed relationship. Such a case could involve

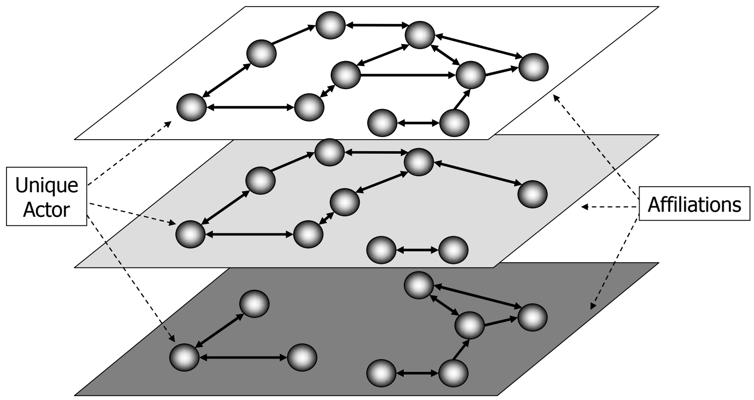


Figure 3.1: Layered Social Network

the difference between a random individual standing beside you on a sidewalk and the lady just introduced to you as the new fiancée of your brother. Both are strangers, yet an inevitable difference between the strengths of the two relationships occurs due to the implied trust gained from a familial context. Consequently, it is suggested that by increasing or at least acknowledging the dimensionality of information gathered on individuals of interest, a better understanding of the overall network behavior can be achieved.

It should be noted that this approach has at least two potential drawbacks. First, methods attempting to measure strength of ties are generally criticized when applied to non-cooperative networks, such as terrorist organizations, in that increasing sophistication of analysis methodologies "may still not yield a more useful map" towards understanding the underlying network behavior [Fellman and Wright, 2003, pg. 5]. Nonetheless, ignoring this type of information automatically presumes all interpersonal ties are homogenous. When considering leaders, followers, and actors that serve as bridges, liaisons, and gatekeepers, knowing not only which particular individual may successfully be exploited, but which of their interpersonal relationships is important when evaluating the efficacy of information operations courses of action.

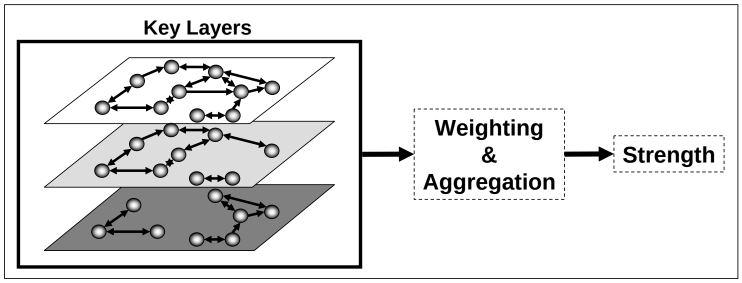


Figure 3.2: Layer Aggregation and Strength

The second potential drawback deals with the natural tendency for individuals to cognitively account for multiple relationship contexts simultaneously. This implies that when attempting to ascertain a flow of information or influence throughout a social network, decomposition-like techniques are likely inappropriate models of the way individuals conduct interpersonal relations and exchanges. However, if a decision maker were interested in leveraging weaknesses or strengths within a specific context so as to disrupt or affect overall network connectivity, strength, and so forth, single-context analyses may lend themselves to such objectives. Both problem aspects, model formulations, and potential applications are discussed in later chapters.

The dilemma then is that the second drawback suggests that multiplexity must be considered as a combined effect between two individuals, whereas the first drawback maintains that it should not. The impasse can be resolved when considering the statement that "[multiple relations should not be combined] unless there is a substantive reason for doing so" [Wasserman and Faust, 1994, pg. 219]. For this research, the 'substantive reason' is the proposition that multiple, interpersonal ties contribute to relationship strength in the manner proposed by Granovetter, suggesting a probably linear combination of contexts [Granovetter, 1973, pg. 1361].

It is hypothesized that, in general, the more individuals have in common, the stronger the relationship between them, as suggested by Haythornthwaite [1999]. Chapter VI explores a number of combination techniques and their implications, some of which involve the allocation of weights for each layer. The weighting corre-

sponds to the importance that the target network in toto places upon a particular relationship context. As the majority of this data is unlikely to be directly measurable, expert opinion familiar with the culture, indoctrination procedures, and institutional foundations will always play a vital role in providing guidance regarding the weights. However, the use of a dynamic weighting can help ascertain the impact of information operations upon the network relationships. For example, compare the bottom layer to the top two in Figure 3.2. If information operations marginalized the weight of the top two layers from the individuals' point of view, a fissure between the network members may be observed. Therefore, despite the ability to measure exactly how much each context contributes to the strength of interpersonal relationships, the sensitivity of a given network to perturbations of the weight set, and the subsequent impact upon associated measures can be explored. As today's terrorist organizations are increasingly multi-cultural, extensions allowing for individual-specific weight sets are also examined.

An additional measurement aspect of this research is the theory of gains and losses. This research effort attempts to extend the work of French [1956] and Friedkin [1986] regarding social power. Couched in the goal of explaining the dynamics of opinions among individuals, both works essentially suggest that "the influence process in groups can be explained in terms of patterns of interpersonal relations" [French, 1956, pg. 81]. French defines interpersonal power as "the power of A over B (with respect to a given opinion) is equal to the maximum force which A can induce on B minus the maximum resisting force which B can mobilize in the opposite direction" [French, 1956, pg. 183].

This research suggests that such interpersonal power, or lack thereof, is analogous to gains or losses. Interestingly, French mentions bases of power that are not necessarily related to network topology as focused upon by both French [1956] and Friedkin [1986]. The bases are attraction, expert, reward, coercive, and legitimate

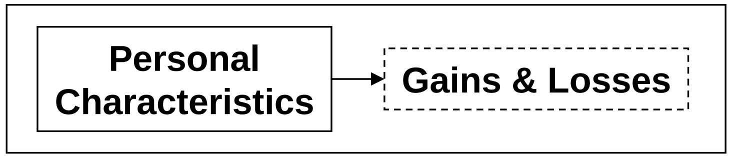


Figure 3.3: Sources of Gains and Losses

power. All of these bases connote factors extending beyond network topology, such as personal characteristics or formal and informal organizational roles.

For example, attraction power may be dependent upon the charisma, appearance, and professional, educational, and religious backgrounds of the individuals. Within the context of a generalized network flow model, any two given individuals perceiving each other as their peer would result in the power of one over the other to be the same, resulting in a multiplier equal to one. Alternatively, an individual may demonstrate a greater influence, pressure, or power over another due to a socioeconomic or status differential, resulting in a multiplier > 1. Lastly, communication emanating from an individual that is perceived as an underling, unreliable, untrustworthy may carry a negative undertone, resulting in a multiplier < 1. However, the only basis French elaborates upon is that of attraction, noting that cohesiveness has been 'operationalized' in past studies to account for attraction between individuals within a network [French, 1956, pg. 185]. Figure 3.3 depicts the general notion of this concept, which is further detailed in Section 3.3.4 and Chapter VII.

It is therefore hypothesized that the combination of tie strength and measurement of gains and losses of influence provides a more robust and capable model of network behavior. Additionally, the act of seeking this information inevitably contributes to an improved understanding of the extent and nature of the interpersonal relationships of the target network. A new social network analysis measure, and new techniques derived from its application, to accomplish this in a more efficient manner, is developed in Chapter IV.

When it comes to affecting the network in some way, presumably via information operations that may involve kinetic and non-kinetic means, the evaluation of potential target sets and courses of action must be accomplished. Chapter V builds upon and extends the analytic capabilities of the key player concept through mathematical programming techniques. Chapter VII combines the measurement of multiplex relations discussed in Chapter VI with the measurement of gains and losses. These new techniques facilitate a generalized network flow model of influence similar to that described by Renfro [2001], which can be used to evaluate influence courses of action.

The overarching goal of this research is to use these combinations of models to improve understanding of potential behavioral patterns that belie the target network, and their reactions to the information operations that may be imposed upon them. Subsequently, such understanding may be used to develop improved courses of action to effectively achieve a specified change in behavior in one or more actors within the target network. These efforts, and the research endeavors presented within this dissertation, require several underlying assumptions.

3.2 Assumptions and Limitations

As discussed in Chapter I, it is assumed that the data required for the methodologies presented is available and known with certainty. The mathematical nature of the approaches presented permit the investigation of relaxations to this assumption via sensitivity analyses, involving both 'one-at-a-time' and parametric approaches to the impact of uncertainty. However, these techniques may not always avail themselves to determining the potential effects when uncertainty pervades the entire network.

Since one of the objectives requires the forced flow of influence through a network, specific paths upon which the influence travels result. As with many social

network analyses, independence is assumed, potentially contrary to typical social interactions. For example, when one person passes on a message, information, or influence to another, that person may also provide the previous source, either for informational purposes or for emphasis. As an example, in reality there may be a different influential impact between "our co-worker said we need to do..." and "our boss said we need to do...." This scenario implies dependence and, despite a potentially closer tie to reality, is assumed not to hold for computational convenience [cf. Friedkin, 1986].

The other major assumption underlying this research is the static approach to network topology, both the structure and the perceived strengths of relationships. It is certain that over time, some individuals may change their opinions or strategies, relationships evolve and devolve, and the overall social network structure changes due to recruitment of new individuals, new opportunities for interaction, and departures from the network. However, given the nature of available intelligence information and the near-term focus to which these techniques are amenable, it is assumed that key changes in the network are primarily due to the actions or influence imposed upon it. With the possible exception of dynamic programming, this appears to be the preferred way to deal with other, albeit open, social networks in current sociological and anthropological literature. Other efforts are pursuing the capability to simulate dynamic network behaviors, Carley [2003], for example; however, this also requires collecting a great deal more information that may or may not be available.

3.3 Approaches

The theoretic and applied contributions from this research involve successive, conceptual steps that may, as applicable, build upon each other or provide complimentary analyses. The steps may either be used stand-alone or combined to provide a theoretically sound analysis methodology for the study of layered social networks. As mentioned earlier, some concepts are more amenable to layered networks than

others. The social networks of particular interest are comprised of adversarial and non-cooperative individuals; however, the proposed methodologies are not necessarily limited to such organizations.

The most benefit gained from the approaches developed occurs when there exists data characterizing both (1) the dimensions of interpersonal relationships and (2) individual attribute data. Although the most detailed level of analyses requires both classes of data elements, some approaches may still be used if only limited data is available. Each of these approaches are generally described in the following sections; detailed discussion and notional analyses of each are presented in following chapters.

3.3.1 Screening

A new social network analysis measure is offered that attempts to assess an actor's position from the standpoint of centrality, power, or prestige. The position type is dependent upon the sociometric data available. Use of symmetric adjacency matrices yield actor centrality. Use of asymmetric adjacency matrices can yield both power and prestige positions. This reach-based measure builds upon concepts such as information attenuation as a function of path length [cf. Katz, 1953; Stephenson and Zelen, 1989] and reach type [Valente and Foreman, 1998]. Although this measure is restricted to the analysis of dichotomous networks, it has several analytic advantages over similar, traditional social network measures. The mathematical development of this measure, its theoretical bases, and the characteristics that make the measure amenable to the study of non-cooperative networks are provided. Methods to apply this measure to layered networks are presented.

3.3.2 Targeting

The appealing concept, from a military perspective, of the KPP-2 problem discussed in Section 2.2.2 is extended in a variety of ways. New mathematical pro-

grams are developed to provide a more robust alternative to the insightful, but initial, heuristic approach devised by Borgatti [2003b]. Assuming that some individuals within a target network are accessible, or can be made accessible, the models and techniques developed in this research offer a means to build and evaluate possible target sets. There are, of course, advantages and disadvantages of this approach.

Advantages over Borgatti's approach include: (1) the ability to address asymmetric networks; (2) the ability to solve to optimality the *kp*-set; (3) the collection of multiple optimal solutions; and, (4) the ability to extend the problem aspects to other dimensions of interest. Relationships between classic operations research models such as covering and partitioning problems, dominating sets, and the *p*-median problem and the sub-objectives of the KPP-2 concept are established. This not only offers improved solutions, but enables improved capabilities through the blending of models and the incorporation of specialized constraints.

Examples of such capabilities include the use of goal programming to evaluate trade-offs between competing objectives (e.g., *kp*-set size and network coverage); the *a priori* designation of individuals as accessible or inaccessible; the incorporation of costs, real or perceived, to access and co-opt a key player; and, the enumeration of multiple optimal solutions, thus providing more viable and optimal alternatives for the decision maker. These methodologies and illustrative examples are provided.

3.3.3 Measuring Multiplexity

There are few sources within the literature to date that deal directly with attempts to measure quantitatively relationships that vary in strength, other than those that, for convenience, assume or hypothesize the researcher could obtain those values. Of those that do acknowledge the dimensionality of interpersonal relationships, none, with the exception of Renfro [2001] and Clark [2005], have thus far been uncovered that attempt to take multiple contexts or characteristics and formally

develop an approach to differentiate quantitatively relationships between pairs of individuals.

Measuring the strength of a relationship by incorporating a number of appropriate contexts comprises the first step and provides inputs for other model formulations. Contexts of interest are drawn from Marsden and Campbell [1984] and Hite [2003] (as discussed in Chapter VI) as starting points. However, depending upon the nature of the intelligence collection process, contextual analyses may only be feasible after an initial set of individuals has been established and a common set of contexts that comprise their relationships have been discovered.

Note also that multiplexity is only one of many aspects that could contribute to the strength of a tie. Previous works such as Granovetter [1973], suggested duration and frequency serve as proxies for multiplexity. Consequently, a model defining the theoretical building blocks that contribute to the 'strength' of a relationship is developed. Other works such as [Carroll, 2006; Friedkin, 1990, 1980; Gould, 1991; Marsden and Campbell, 1984] are analyzed, summarized, and synthesized to serve as a basis for the theoretical model. This effectively builds upon the seminal work of Granovetter, who stated the general definition of the strength of an interpersonal tie is "...a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" [Granovetter, 1973, pg. 1361]. Areas left 'for future study' by Granovetter include the "operational measures and weights attaching to each of the four elements..." [Granovetter, 1973, pg. 1361]. Such weighting schemes that accompany the model of tie strength are discussed in Chapter VI.

Renfro [2001] evaluated pair-wise comparisons of individual psychological states to determine asymmetric social closeness values. Clark [2005] developed a weighted combination of structure-based, pair-wise centrality results that were further weighted by a method accounting for individual characteristics. The article by Friedkin discusses the construction of a Guttman scale–where different stages or assessments can

imply others–that incorporates claims of frequent discussion, of seeking help, and of close friendship [Friedkin, 1990]. Friedkin's work, however, assumes that each of these claims contributes equally to the tie strength. This may ultimately be a consequence of the Guttman scale approach. Other than these, none of the preliminary works mentioned directly address this concept in any reasonable detail, especially if an analyst wanted to associate a value indicating strength with an edge in a social network graph using different types or classifications of network data (e.g., familial, training, and so forth).

Therefore, this research effort incorporates the use of decision analytic techniques to capture the essence of the model initially proposed by Granovetter, while accounting for the taxonomies provided by Hite and the findings that may have changed initial perceptions of what makes a strong tie; for example, Marsden and Campbell found that "there are difficulties with frequency and duration of contact as indicators of strength" [Marsden and Campbell, 1984, pg. 482].

It is presumed that such a contextual aggregation technique involves weights that, from the adversary's perspective, indicate the importance a given context plays within interpersonal relationships [cf. Clark, 2005]. In addition, if psychological operations are applied to one or more layers, but not necessarily all of them, investigation of how these weights may change over time and the affect upon the network performance and exchange of influence (or power, or status, etc.) in response to these external forces–courses of action–are performed. Determining the contexts or layers of interest is potentially one of the more difficult areas of this research, as the types of ties that result in the strong, trusting relationships are likely dependent upon the origins of the organization and the scenario under analysis, or simply predicated upon the available intelligence information.

Regardless of the technique chosen to perform future analyses, a key aspect that must be considered in this area is that of data acquisition, particularly when dealing with non-cooperative networks. Consequently, a secondary goal of this research,

regardless of the area, is to develop the theory and requisite methodologies amenable to information that must be gathered from a distance.

All of these efforts serve as an ideal input for an extension of Renfro's flow model formulation from a single, possibly multi-commodity network to a multi-layered network is planned. Although Renfro proposed a multi-commodity flow formulation, characterization and analysis of non-cooperative networks as layered, inter-dependent network formulation is investigated in greater detail.

3.3.4 A Generalized Social Network Model

The overall goal of this research thread is to quantify an individual's power of persuasion over another. Renfro suggested that gains and losses within a generalized network model were analogous to preconceived notions or poor communication. This research contends that these effects are due to a person's ability, or lack thereof, to persuade other individuals. Ultimately, this research effort relies upon works such as persuasion theory [Seiter and Gass, 2004]; methods incorporating individual attribute data [Clark, 2005]; and, the bases of interpersonal power [French, 1956].

Two immediate benefits are derived from a measure of interpersonal persuasiveness. First, this measure is incorporated into a generalized network flow model of a social network. This provides a means to examine the efficacy of potential courses of action in further detail. These courses of action, in general, seek to influence specific, and presumably inaccessible, actors within a target network. The influence is indirectly transferred by using accessible actors and the target network's own relations. Analysis methods and techniques taking advantage of sensitivity analysis procedures are discussed with the objective of improving insight into the adversary.

Another immediate use for the gains and losses measure–as well as the measurement of relationship strength discussed earlier–is the incorporation into information flow centrality measures as presented by Freeman et al. [1991] and Brandes and Fleischer [2005]. For example, Freeman et al. [1991] developed a centrality measure

by evaluating various maximum flow characteristics of the network. The flow model required arc values, representing the strength of a relationship, which were assumed to be available. Elaborating upon these 'strength' values via the aggregation of contextual layers benefits this research area in a fashion similar to that of Renfro [2001] and Clark [2005]. However, the inclusion of gains and losses due to persuasion also provides an extension of Freeman et al.'s centrality measure, incorporating a generalized, and assumed to be more representative, network flow model. Justifications for inclusion of this aspect of interpersonal behavior are discussed and, as accomplished by Freeman et al. [1991], comparisons to the initial network flow centrality measure and other measures of betweenness is explored.

All of these measures and methodologies are expected to provide information and insight regarding actors of interest or the implications of imposing external influences upon a target network. The assumptions and flow processes underlying network flow models of social networks, however, must be carefully considered.

3.3.5 Analysis of Layered Social Networks

Verification, comparisons to other techniques, and demonstrative analyses cannot be performed without data. Subsequently, notional networks are evaluated throughout the document, illustrating the various aspects of this research. The application of all techniques developed within this dissertation to a subset of the Al-Qaeda terrorist network is presented in Chapter VIII. This case study demonstrates the various theories and associated methodologies developed within this dissertation.

3.3.6 Summary

The overall framework for this research effort is depicted in Figure 3.4. The flow of research tasks is from left to right; the arrows indicate predecessors and relations between activities. This framework not only serves as an overall analysis

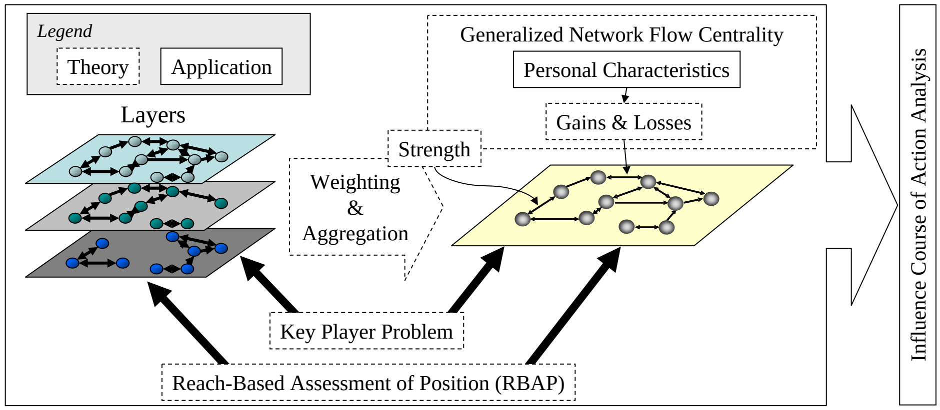


Figure 3.4: Research Framework

methodology, but also helps to identify and tie together the theoretical aspects of this research.

This framework, composed of an interrelated and complimentary suite of analysis approaches, facilitates the gaining of insight into the adversarial network. Such insight can be leveraged in a number of ways. For example, the development and estimation of the efficacy of courses of action against the network and improved capabilities to forecast roles and responsibilities of individuals within a non-cooperative network–all despite limited information. The next chapters provide the details of the research approaches that have been, very generally, described.

IV. Screening

4.1 Chapter Overview

A number of centrality measures that rely upon the structural characteristics of a social network to assess the importance of each actor within the network exist [cf. Wasserman and Faust, 1994]. In the context of evaluating non-cooperative, clandestine networks, measures that can be efficiently calculated and perform well despite limited information are of increasing interest to counter-terrorist applications of social network analysis.

The measure developed in this chapter is designed to serve as a screening tool to identify individuals within a given adversarial, clandestine network (an active terrorist organization, for example) who may play important roles in achieving organizational objectives. Those actors with such roles are of interest and are candidates for focusing intelligence surveillance and analysis resources. *Importance* is assumed to be positively correlated with how easily or efficiently a given actor can communicate, directly or indirectly, with all other actors in the group.

In the context of typical network analyses, the nature of such roles is often predicated upon network topology. For example, network data that captures directed relationships invokes the notions of prestige and power. A prestigious actor is "one who is the object of extensive ties" [Wasserman and Faust, 1994, pg. 174]. Alternatively, a powerful actor is one that "influences the behavior (either overtly or covertly) of others in accordance with his own intentions" [Goldhamer and Shils, 1939, pg. 171]. Power thereby implies a focus upon ties emanated. Symmetric (undirected) data simply fall within the study of, and have accompanying measures to assess, actor centrality [cf. Wasserman and Faust, 1994, Chp. 5]. This measure is shown to be applicable to all three of these analytic categories (prestige, power, and centrality), assuming that appropriate steps are taken to ensure that the measure is indeed capturing the information or influence flow of interest [cf. Borgatti, 2005].

Reach-based assessment of position (RBAP) was initially motivated by the concepts within the status measure of Katz [1953], discussed in Chapter II. This effort ultimately resulted in computational and theoretical changes making Katz's measure potentially more suitable for analysis of clandestine networks that rely heavily upon secrecy for their operational success [cf. Post, 2005; Baker and Faulkner, 1993]. In addition, RBAP is conceptually related to the *radiality* measure developed by Valente and Foreman [1998], which "refers to the degree an individual's relations reach out into the network providing access to many and diverse others" [Valente and Foreman, 1998, pg. 90].

This initial development of RBAP is focused upon screening a clandestine network, characterized by a binary (i.e., denoted by 1 if there exists a relationship between two individuals, 0 otherwise) and not necessarily symmetric adjacency matrix, for actors with the most power or influence over all other actors. Thus, the measure can be viewed as (1) a variant of out-degree centrality [Wasserman and Faust, 1994, pg. 199]; (2) a modification of the status measure by Katz [1953]; and, (3) a modification of eigenvalue-based centrality measures that are similar to Katz's measure [Bonacich, 1987; Bonacich and Lloyd, 2001, pg. 195]. The screening process attempts to identify the most interesting actors by virtue of very high, or very low, RBAP scores. This group of actors would then serve as the focus of limited intelligence resources.

In the following sections, the aspects and criticisms of related works are briefly reviewed, and the theory underlying RBAP is developed. The chapter concludes with a demonstrative example and discussion on how this measure may be applied to layered social networks.

4.2 Background

Given a dichotomous representation of an adversarial, clandestine network, a measure that seeks to identify actors that are able to reach or influence all other actors within the network as quickly as possible is desired. In the context of a dichotomous network, quick refers to the number of steps between actors within a network. This automatically invokes a common underlying assumption prevalent in many social network analysis measures–that influence or information propagates through a network via shortest, or geodesic, paths. Recall that the geodesic path is defined as "the (not necessarily unique) shortest path through the network from one vertex to another" [Newman, 2003, pg. 173]. The definition, however, implies that there could be multiple shortest paths of a given distance between any two given actors, a phenomena leveraged in the classic betweenness centrality measure as well as RBAP [Wasserman and Faust, 1994, pg. 188-91].

From a communications point of view, flow via the shortest path may minimize the likelihood and impact of errors or misperceptions that often plague human interaction. However, as several authors have contended, communication or influence between individuals within a clandestine network may not necessarily flow along the shortest path. For example, regarding the impetus behind their centrality measure that accounts for all possible paths between any two individuals, Stephenson and Zelen [1989] point out,

> it is quite possible that information will take a more circuitous route either by random communication or may be intentionally channeled through many intermediaries in order to 'hide' or 'shield' information in a way not captured by geodesic paths [Stephenson and Zelen, 1989, pg. 3].

Other previous works suggest that when an organization is faced with tradeoffs between efficiency and concealment, the subsequent network structure evolves in a manner that may be contrary to classical sociological expectations [Krebs, 2002; Baker and Faulkner, 1993, pg. 856]. However, the actual interpersonal communica-

tion may still follow the shortest path relative to the secretive network, despite the fact that such a path could be shorter if the network were operating and evolving freely without recourse. If secure communications are required, one can assume that longer communication chains offer more opportunity for interception of message traffic and associated operational risk. Hence, communication among paths other than the geodesics is (potentially) contrary to the organizational goals of secrecy [e.g., Post, 2005, Chapter 2].

The following sections discuss existing measures that lead to the development of RBAP. These measures include the status index of Katz [1953], the radiality and integration measures proposed by Valente and Foreman [1998], and the centrality measures for asymmetric relations developed by Bonacich [1987]; Bonacich and Lloyd [2001]. For a more comprehensive comparison between the related measures, the reader is referred to Wasserman and Faust [1994, pg. 198-219] and Borgatti and Everett [2006].

4.2.1 Contributing Measures

Recall the discussion of the status measure proposed by Katz in Section 2.4.1. Katz's recursive status measure, taking advantage of the convergence of a geometric series, captured 'all possible walks' of infinite length with a relatively easy calculation and the reasonable assumption that the effect of communication or influence along a path decreased as a function of the path's length. However, this measure suffers several conceptual and theoretical problems, particularly when considering and analyzing the behavior of non-cooperative social networks. These problems include: the characteristics of the flow assumed and actually captured by the calculations; the potential length of the paths implicitly accounted for within the measure's calculations; and, the arbitrary choice of the attenuation factor.

As mentioned in Section 2.4.1, the flows captured by Katz's index include not only all possible walks, but all possible paths, as well as directed edge sequences

that fit neither the walk nor the path definition–an observation previously noted by [Leenders, 2002, pg. 32]. To further complicate the potential application of this measure to non-cooperative networks, directed edge sequences of infinite length are also incorporated within the status values. These sequences do indeed contribute to the status scores of the individuals and effectively measure phenomenon that is contrary to the behavior inherent within the networks of interest. Recall Theorem 1, due to Deo [1974], which categorizes the entries within the powers of the sociomatrix as either directed paths from i to j, directed walks from i to j, or directed edge sequences that are neither paths nor walks.

Directed edge sequences that fall within the second category, walks, could be perceived as inefficient communications practices among network members. For example, suppose actor A wanted to transmit a message or influence to actor D. With other actors B and C, a valid walk, assuming the network connectivity exists, from A to D could include A-B-C-A-D. In this particular example, it is more efficient for actor A to communicate directly with D instead of routing the same message through B and C only to have it return to actor A.

Similarly, directed edge sequences that fall within the third category essentially include repetitive banter between two or more individuals. For example, with two actors A and B, a directed edge sequence of length 3 ($p = 3$) could include the path A-B-A-B. With three actors A, B, and C, directed edge sequence of length 5 ($p = 5$) between A and C could include the path A-B-A-B-A-C. Assuming that each interaction or period of communication imparts a potential risk of being uncovered, captured, or providing an adversary with additional and sensitive information, Katz's status measure is inappropriate for analysis of non-cooperative or clandestine networks.

In addition, the potential length of walks measured goes to infinity. Again, this would involve an infinite amount of communication exchanges between the individuals, which would likely be counter to operational security objectives. Note also

that, given a graph with n vertices, the maximum length of a path is $n - 1$. The maximum length of a walk given the same graph and assuming that the path A-B is considered different from the path B-A is $n(n-1)$. Unless the network of interest is an infinite graph, Katz's status index provides an unrealistic characterization of information or influence patterns within an organization. This suggests that a more direct, path-based approach, limited to the length of a worst-case scenario–a path of length $(n - 1)$.

The final points of contention include the arbitrary nature of the attenuation factor, α, as well as the restricted range of its acceptable values being predicated upon the network structure. These facts detract from the overall analytic power of this concept and resulting measure. Even within the ranges recommended by Katz, the most 'central' actor is often dependent upon the value of α; explaining this phenomenon is even more complicated when dealing with an infinite number of paths, walks, and directed sequences levied upon a finite graph.

As an example, using the notional network in Figure 2.8, Figure 4.1 depicts the results of Katz's measure with varying levels of α within the recommended range. The graph simply captures the rank order of the status for each of the six actors, with the values 6 and 1 indicating the highest and lowest ranking status scores, respectively. Interestingly, two crossover points exist, resulting in actors A and D exchanging status rankings around $\alpha = 0.36$, and actors A and F exchanging status rankings at around $\alpha = 0.48$. If the value of α is likened to the attenuation of influence or status as a function of distance, the range, and therefore the possible range of assumptions regarding the amount of attenuation, is unfortunately limited by necessity in order to determine a solution to the system of equations.

Related approaches, based upon the eigenvectors of $\mathbf{X}$, due to Bonacich and Lloyd [2001] and Bonacich et al. [2004] are frequently used within the sociological literature. Recall from Section 2.4.4 that eigenvector centrality views the nature of power or status from recursive standpoint, thus sharing similar conceptual and

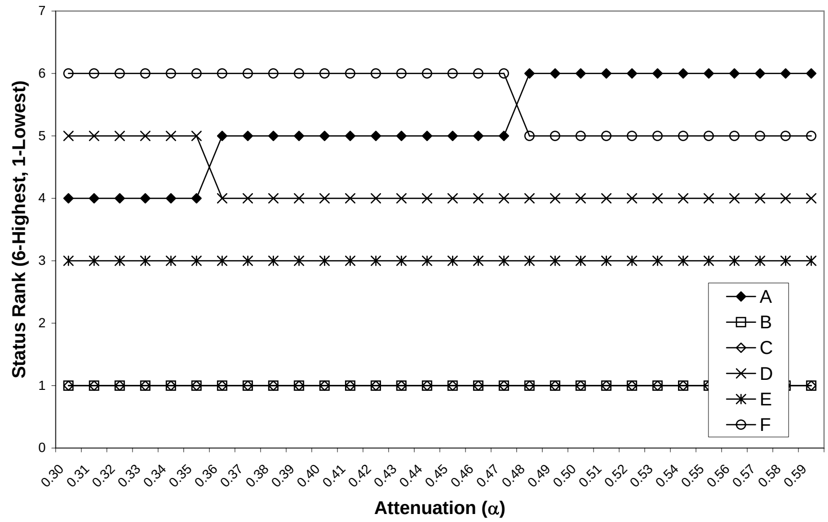


Figure 4.1: Change in status with attenuation

mathematical underpinnings to the measure developed by Katz. Consequently, this measurement approach suffers similar difficulties regarding the types and lengths of flows of information, status, or influence between a network's individuals. Although there is an α component of eigenvector centrality, it is used to tradeoff the importance of exogenous (e.g., actor attributes) and endogenous (i.e., network topology) factors as contributing to status. Ultimately, the value and permissible range of this particular parameter is also predicated solely upon network structure.

Valente and Foreman [1998] developed a dual-purpose measure based upon a reverse geodesic distance approach. Given that the measures of interest are integration ("can be reached by many others rapidly") and radiality ("the degree to which an individual's relations reach out into the network"), the measure is dual-purpose in that the input is either the adjacency matrix or its transpose, respectively [Valente and Foreman, 1998, pg. 90]. The integration measure for a given actor k is formally defined by

$$I\left(k\right) = \frac{\sum\limits_{j \neq k} RD_{jk}}{n-1}, \tag{4.1}$$

where n is the total number of actors within the network, RD_{jk} is the reverse geodesic distance, computed by subtracting the geodesic distance between j to k from $(1+D)$ [Valente and Foreman, 1998, pg. 92].

To calculate *radiality*, the measure is applied to the transpose of the adjacency (or nomination) matrix [Valente and Foreman, 1998, pg. 93]. Note that there is no attenuation factor associated with longer geodesic paths. In addition, RD_{jk} only counts single instances, if they exist, of a geodesic path between any two given actors. Consequently, radiality may not truly capture "the degree to which an individual's relations reach out into the network" if multiple shortest paths implies more potential for the exertion of influence or power.

Lastly, a reach-based measure of centrality that "counts the number of nodes each node can reach in k or less steps" is offered by Borgatti et al. [2002]. This too can be applied to directed and undirected networks, as well as give indications of status or power. However, this particular measure does not accommodate multiple shortest paths and, from the documentation available, the method of attenuation, if any, is neither immediately apparent nor theoretically justified.

Although the measure developed by Katz is easy to implement, Katz's measure captures network behavior that goes beyond the circuitous paths posited by Stephenson and Zelen. Therefore, a new measure is sought that assesses the potential influence an individual can propagate throughout the network via efficient information channels (i.e., shortest paths), which also accounts for the number of options available for information flow via multiple shortest paths. From these previous efforts, an opportunity clearly exists to (1) enhance the concept of radiality and integration posited by Valente and Foreman [1998] and (2) separate the concept of 'attenuation' from the conditions required for system solution.

4.3 Assumptions and Development

What differentiates RBAP from existing measures of power is (1) the use of multiple instances of shortest paths; (2) the process of accounting for the options available, if any, to the actors regarding alternative shortest paths; and, (3) uncoupling the concept of 'attenuation' from conditions necessary for a system solution.

Recall that the shortest path between any two individuals of a connected network ranges between 1 and $(n-1)$. Deo's theorem is extended via Corollary IV.1 to enumerate the number of all pair-wise shortest paths, by raising the adjacency matrix to powers ranging from 2 to $(n-1)$. Note that this approach requires a dichotomous representation of the network; therefore, this measure assumes that all links are of length 1.

Corollary IV.1. *Given an adjacency matrix* $\mathbf{X}$*, by raising it to the power* p*,* $p = (1, \ldots, n-1)$*, the first non-zero element* $x_{ij}^p, i \neq j$ *of* $\mathbf{X}^p$ *yields the number of shortest paths of length* p *from* i *to* j*.*

Proof. $\forall\, x_{ij}^p \in \mathbf{X}^p$, $i \neq j$, $x_{ij}^p > 0$, and $x_{ij}^k = 0$, $k = 1, \ldots, (p-1) \Rightarrow$ no directed edge sequences of length $1, \ldots, (p-1)$ exist. Therefore, the shortest path between i and j must be of length p and further implying that the value x_{ij}^p must fall in the first category stated by Deo, which is the number of directed (shortest) paths from i to j. $\square$

Use of Corollary IV.1 facilitates the enumeration of shortest paths and their lengths between all actors. From this, the following definitions serve as the basis for RBAP.

$\alpha \triangleq$ an attenuation factor, with a similar, penalizing purpose to that used in [Katz, 1953]; however, for RBAP there is no restriction other than $\alpha \in [0, 1]$

$R_{\mathbf{X}^i} \triangleq$ $(n \times n)$ matrix that stores the number of shortest paths of length i from any two given actors where the criteria of Corollary IV.1 are satisfied

$r(k)_i \triangleq$ number of other actors reached by actor k via a shortest path of length i

$\mathbf{r}_i \triangleq$ $(n \times n)$ diagonal matrix where $\forall\, r\,(m)_i > 0$, $\mathbf{r}_i(m, m) = r\,(m)_i^{-1}$ for $m = 1, \ldots, n$; zero otherwise

One other underlying assumption of this measure is that the highest level of power is obtained when an actor is adjacent to all other actors within one step. Consequently, the numbers provided in the matrices ($R_{\mathbf{X}^i}$) must be normalized to avoid actors with numerous but indirect paths to all other actors scoring higher than actors that can reach all other $(n-1)$ actors within one step. This process is accomplished with the variable, $r(k)_i$.

For example, consider the network in Figure 4.2. Actor i, reaches three other actors via a shortest path of length 1. Therefore, to reach any other actor, j, the maximum number of shortest paths of length 2 is bounded above by $r(i)_1 = 3$. If the three dashed paths existed in the network, the value of $R_{\mathbf{X}^2}(i, j)$ would be 3. This value and all other values in the ith row of $R_{\mathbf{X}^2}$ are normalized by dividing by $r(i)_1 = 3$.

Suppose further that from node i, two new nodes were reached via a shortest path of length 2 (nodes d and e in Figure 4.3). Therefore, to reach any node j via a shortest path of length 3, there are at most $3 \times 2 = 6$ *possibilities*, given by the paths $(i - a - d - j)$, $(i - a - e - j)$, $(i - b - d - j)$, $(i - b - e - j)$, $(i - c - d - j)$, and $(i - c - e - j)$. Consequently, this requires that the value $R_{\mathbf{X}^{\mathbf{2}}}(i, j)$, as well as all other values in the ith row of $R_{\mathbf{X}^{\mathbf{2}}}$, be divided by $r(i)_1 \times r(i)_2 = 6$. To facilitate this

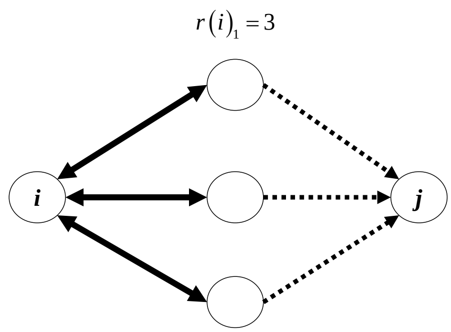


Figure 4.2: Paths to j given $r(i)_1 = 3$

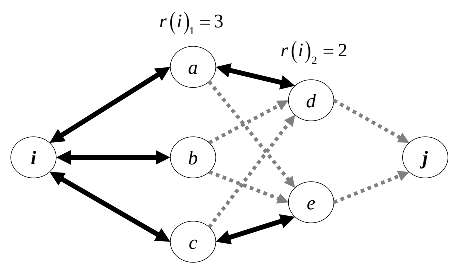


Figure 4.3: Paths to j given $r(i)_1 = 3$ and $r(i)_2 = 2$

calculation, the matrix satisfying the conditions of Corollary IV.1 is pre-multiplied by the matrix $\mathbf{r}_i$.

An attenuation factor, $\alpha \in [0, 1]$, not unlike those seen in related works, represents the diminishing effectiveness of communication or influence as a function of path length. However, unlike the works of Katz [1953] or Bonacich and Lloyd [2001], calculating the RBAP measure is not predicated upon finding a specific value for α. While this does not resolve the ambiguity issue regarding the effects of longer path lengths upon power or status, it does offer some analytical freedom, as α can take on any value within its range without negating the measure's results. Additionally, the attenuation is assumed not to begin until $p \geq 2$. Therefore, RBAP simply reduces to degree centrality (simple, in-, or out-degree depending upon the data and application) when $\alpha = 0$ and is bounded above by the total number of other nodes that can be reached from any given node when $\alpha = 1$. Hence, both ends of the

range offer both conceptually and mathematically meaningful interpretations. The range from 0 to 1 can also be interpreted as an actor's position relative to all others from a local to global perspective, respectively. Additionally, α may be varied in order to perform sensitivity analyses, potentially gaining insight into the positions and possible roles various actors serve within their organization. With $\mathbf{1}$ being an $(n \times 1)$ vector of ones, the $(n \times 1)$ RBAP result is

$$\mathbf{RBAP} = [R_{\mathbf{X}} + \alpha \mathbf{r}_1 R_{\mathbf{X}^2} + \alpha^2 \mathbf{r}_1 \mathbf{r}_2 R_{\mathbf{X}^3} + \ldots + \alpha^{n-2} (\prod_{i=1}^{n-2} \mathbf{r}_i) R_{\mathbf{X}^{n-1}}] \mathbf{1}. \tag{4.2}$$

A proxy measure for α could include the clustering coefficient of the network, denoted $\gamma(G)$, which is the average of the clustering coefficients for each of the actors within a network. The clustering coefficient for a given actor i is denoted $\gamma_i(i)$. Given the number of neighbors of i (b_i), the individual-specific clustering coefficient is the "ratio of actually existing connections between the b_i neighbors and the maximal number of such connections possible $(b_i^2 - b_i)$" [Sporns, 2002, pg. 178] [cf. Watts, 1999, pg. 32-3]. Consequently, higher clustering coefficients may imply more cohesive and interactive groups and therefore lower communication or influence losses (i.e., higher values of α).

Although not a necessary condition to perform the calculations, application of this procedure assumes that the network of interest is connected. Considering that this measure is reach-based, the centrality calculated for isolates is 0, as expected. However, if the graph is comprised of more than one component, all output will be relative to the specific components and not to the network in total. Subsequently, caution must be taken to avoid misinterpretation of the output by unknowingly comparing results among two or more components rather than across all actors, particularly if the values are normalized. If the graph is comprised of several components, analysis should be accomplished on the component(s) of interest, rather than applying this measure to a number of disconnected components at once.

Table 4.1: Katz and RBAP Comparison ($\alpha = 0.5$)

	High	Rank (value)			Low
Katz	A (0.47)	F (0.45)	D (0.41)	E (0.22)	B and C tie (0.04)
RBAP	F (4.25)	D (3.50)	A (3.25)	E (2.37)	B and C tie (1.00)
Katz (mod)	F (0.25)	A (0.24)	D (0.22)	E (0.11)	B and C tie (0.04)

Finally, since the RBAP value for any given actor is bounded above by $(n-1)$ regardless of α, this measure may be normalized for a given network using

$$\mathbf{RBAP}' = \frac{\mathbf{RBAP}}{n-1}. \tag{4.3}$$

Without normalization, the interpretation of RBAP is the number of other actors that can be effectively communicated with, persuaded, influenced, and so forth, ranging in value between an actor's immediate contacts, to the entire network of individuals. With normalization, the interpretation is similar, but is in the context of percent of the other $(n-1)$ actors. Some examples to explore the resulting nature and meaning of this measure are now discussed.

4.4 Discussion

As an initial investigation, RBAP was applied to the transpose of the choice matrix specified by Katz [1953]. This permits a comparison between Katz's status results and the status (as opposed to power) use of RBAP. With $\alpha = 0.5$, the Katz and RBAP status rankings are shown in the first two rows of Table 4.4.

Observing that there are similarities, and differences, between the two approaches, a more equitable comparison was sought between the two methods. Recall that in Equation 2.1, Katz allowed infinite path lengths. Considering that, in the context of influence or communication among clandestine networks, this may be an unrealistic assumption, suppose a limit identical to that imposed for RBAP $(n-1)$ (while still normalizing by the original definition of m) is applied to the Katz mea-

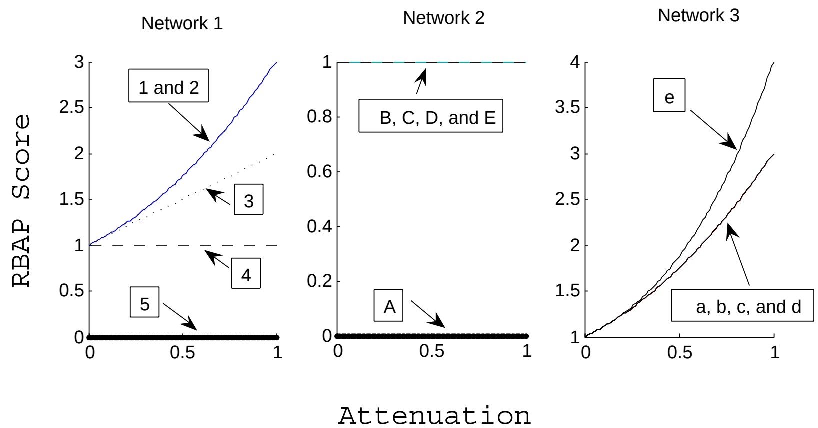


Figure 4.4: RBAP applied to Bonacich and Lloyd [2001] networks

sure.

$$\mathbf{s}_{\text{mod}} = \left(\tfrac{1}{m}\right) \mathbf{1}^{(1\times n)} \sum_{i=1}^{n-1} \alpha^i \mathbf{X}^i. \tag{4.4}$$

The results, denoted 'Katz (mod)' in Table 4.4, show improved comparisons between the two approaches. Spearman's coefficient of rank correlation between Katz and RBAP and between Katz (mod) and RBAP are 0.83 and 0.94, respectively; both are statistically significant at the $\alpha = 0.05$ level of confidence [cf. Lind et al., 2002, pg. 605]. The differences are essentially due to Katz's inclusion of directed edge sequences other than shortest paths. However, given the underlying differences between the measures, perfect correlation between RBAP and any other existing, path-based measure is not one of the research objectives.

Applying the sensitivity analysis procedure for RBAP to all three hypothetical networks discussed by Bonacich and Lloyd [2001] (from Figure 2.9) the results are shown in Figure 4.4 and are as expected. For example, from the perspective of radiality, actors 1 and 2 in Network 1 are more effective than all others in reaching out to the remaining actors. Whereas, actor 5 has no outward connections and therefore has no capacity to influence others. Note that the original purpose for RBAP was to determine the potentially most influential individuals; the results should therefore

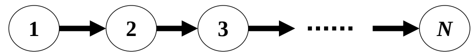


Figure 4.5: Line graph of size N

be considered from the perspective of power. However, to apply this measure with similar objectives as Bonacich and Lloyd [2001] and Valente and Foreman [1998], performing the RBAP measure on the transpose of the adjacency matrix for the same network yields insight from a perspective opposite to power, status. The results generally agree with that of Bonacich and Lloyd [2001], where "unchosen individuals are ignored and have no effect on the status of others" [Bonacich and Lloyd, 2001].

A logical concern for the RBAP is that of computational efficiency. From Equation 4.2, the time required for calculation is dominated by the term, $R_{\mathbf{X}^i}$, which is worst-case $O(N^3)$. Since the measure calculations are complete when all actors have been reached, the worst-case times required for evaluating a given network are dependent upon the network's diameter. To quantify this characteristic, RBAP was applied to a number of line graphs (as shown in Figure 4.5), ranging in size from $N = 10 \ldots 1330$, so that the measure must continue to the largest diameter possible, $(N - 1)$. The performance (in seconds) is compared to N in Figure 4.6.[1] The solid line in Figure 4.6 represents the polynomial (of degree three) equation fit to the data; this can be used as a rough estimate of the worst-case time required to compute the RBAP measure given a network of size N. Noting that the polynomial is increasing substantially with N, and that the size of clandestine networks, particularly terrorist networks, can be much larger than 1330 individuals, worst-case run times may be prohibitive. This limitation is also unfortunately shared by other social network analysis approaches, which use $O(N^3)$ algorithms to determine related measures, such as all-pairs shortest paths and reachability [e.g., Cyram, 2004].

[1]Machine used: Pentium 4, 3.4 GHz, 1GB RAM, running Windows XP Pro.

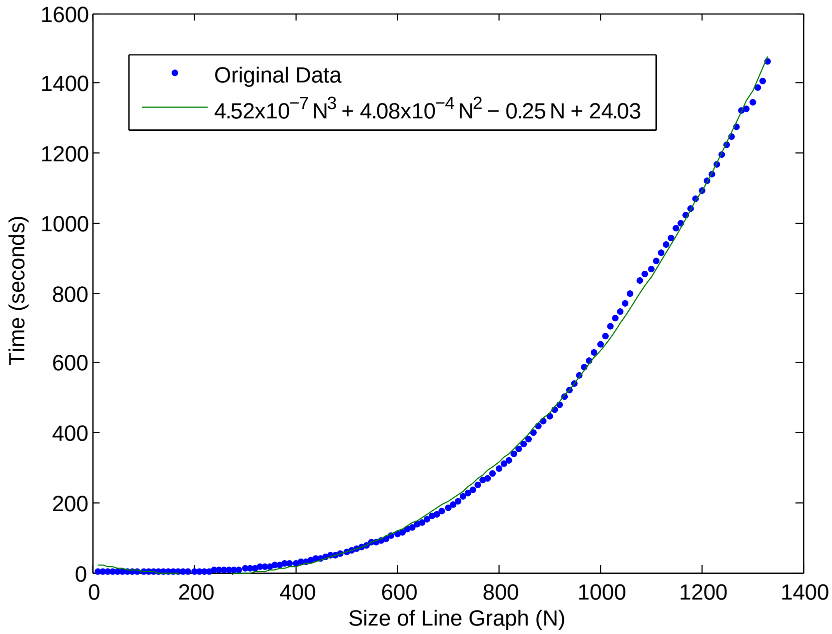


Figure 4.6: Network Size (N) versus RBAP runtime

However, the line graph represents an extreme, and unlikely, topology of a social network, even if the members are engaged in clandestine activity. As an example, the trusted prior contacts of the 9-11 hijacker network analyzed by Krebs had 19 (known) individuals; the diameter of this network, based upon the relationships ascertained from open source data, was 9 [Krebs, 2002, pg. 46]. The relationship between population size and network diameter has been of interest since Milgram traced correspondence paths, wherein the famous six degrees of separation between ostensibly distant and unconnected actors was observed [Milgram, 1967]. Such a six-degree graph would yield a variation of the polynomial in Figure 4.6, and would result in dramatically reduced computational requirements as illustrated in Figure 4.7. Related works have popularized this small-world property [Barabási, 2002; Buchanan, 2002; Watts, 1999].

Numerous connections between real-world, emergent networks and small-world network behavior have been made. Examples include cellular metabolism, Holly-

wood movie-stars, Internet connections and world-wide-web page links, protein regulatory networks within cells, research collaborations, social networks, and sexual relationships [Buchanan, 2002; Barabási and Bonabeau, 2003, pg. 54]. As a result of the small-world property, "their diameter is $O(\log N)$ instead of $O(N)$" [Eppstein and Wang, 2004, pg. 40]. Similar findings have been made in analyzing networks evolving via preferential-attachment mechanisms described by Barabási and Albert [1999] [Liben-Nowell, 2005, pg. 16-8]. In addition, more recent research by Leskovec et al. [2005] has shown diameter to actually decrease with increased network size. These observations translate directly to corresponding savings in RBAP computational performance. Figure 4.7 summarizes the run time required to perform the RBAP measure for networks ranging from 100 to 1400 nodes with varying diameters as opposed to the worst-case diameter of $(N-1)$. As expected, if $D \ll N$ then the computation time required is reduced significantly. For example, the 1300-actor network with $D = 1299$ required 1344.9 seconds to complete. A comparable 1300-actor network with $D = 30$ required 38.8 seconds. Therefore, in lieu of real-world, large, terrorist network data sets, initial experimental results indicate that this is a promising approach with regards to computational efficiency.

The equivalence between social networks and network data gathered to characterize actors and relationships enmeshed within clandestine activity remains an open question. The object of study is still comprised of people with links indicating some form of interaction. Fortunately, previous authors have addressed some of the issues that often plague the application of social network analysis techniques to clandestine networks. For example, several efforts have studied the implications of network sampling upon classic centrality measures using social network data Costenbader and Valente [2003] and random networks Borgatti et al. [2006]; the former concluding that the stability of measures is dependent upon network topology, and the latter indicating stability, using random graphs, is somewhat predictable, particularly for denser networks. In addition, there is increasing interest in applying social network

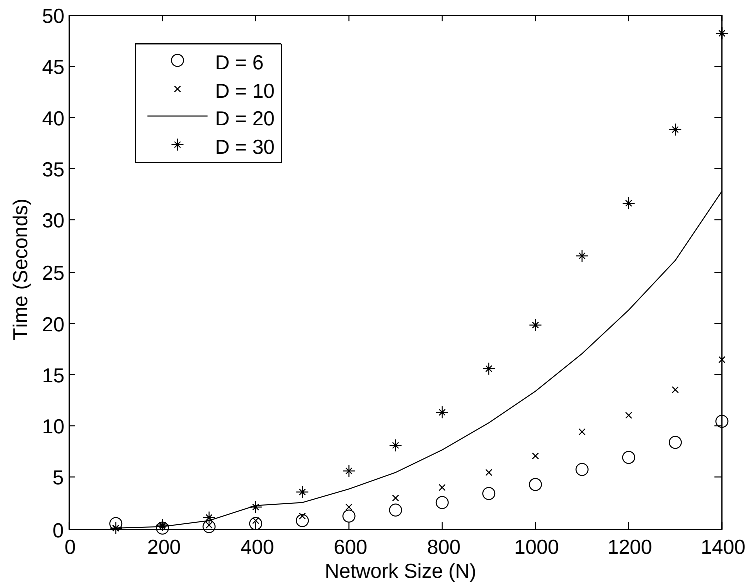


Figure 4.7: Impact of Diameter upon RBAP runtime

analysis techniques to terrorist organizations [cf. Krebs, 2002; Carpenter et al., 2002; Carley et al., 2002; Fellman and Wright, 2003; Thomason et al., 2004, to name a few] Consequently, for this research, it is assumed that clandestine networks are indeed social in nature and will ultimately exhibit the small-world property such that the diameter (D) will be much less than the number of actors within it (N). Since the practical computational bounds of RBAP are dependent on D, this property alone will contribute to improved performance, given reasonably-sized networks.

Given that the underlying motivation for this measure is to provide a means to identify potential actors within an adversarial network that exhibit greater amounts of power or influence among the others (i.e., leaders, potential leaders, coordinators, liaisons, etc.), an analysis of the hijacker network presented by Krebs [2002] is of interest. Analysts must always consider that the adversarial network is constantly trying to either avoid detection or steer our resources in their favor [cf. Sparrow, 1991; Xu and Hsinchun, 2004; Baumes et al., 2004].

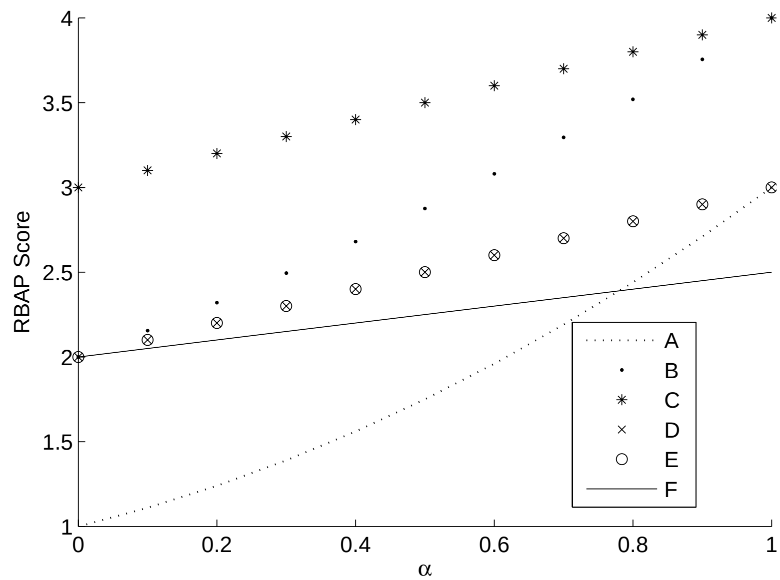


Figure 4.8: RBAP and Katz Network

4.5 Examples

Applying the RBAP measure over the range of α to the Katz network shown in Figure 2.8. The results in Figure 4.8 show similar behavior to Katz's measure in that the most influential individual is dependent upon the level of attenuation selected. However, for RBAP, all values of α provide valid results, given that the attenuation level is justified by careful analysis of the network as a whole. Note that at $\alpha = 0$, the RBAP measure reverts to simple out-degree centrality. At $\alpha = 1$, the RBAP scores are bounded above by the number of reachable actors. The most influential, actors B and C are able to reach all other actors but have limited options in doing so. Such topological consequences are captured by $\mathbf{r}_i$ and therefore results in scores less than $(N - 1 = 5)$ for these actors.

Turning now to the 9-11 hijacker network studied by Krebs [2002], the network of trusted prior contacts was extracted from open source information and is shown in Figure 4.9. To facilitate analysis an identifying number, shown in parenthesis by each hijacker's name, was added. Note also that the resulting graph is undirected;

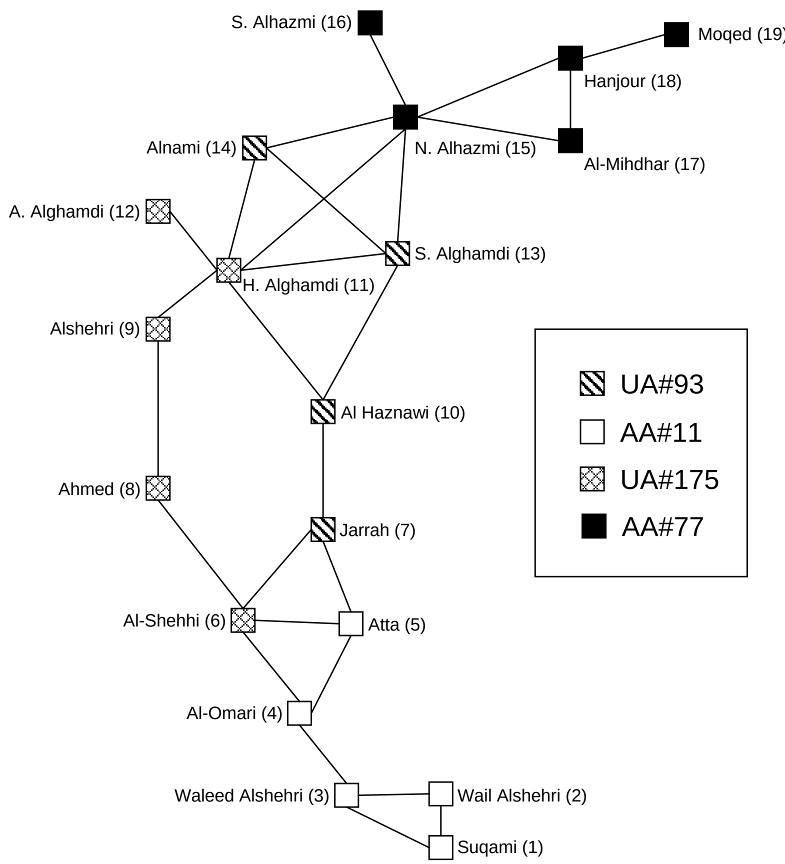


Figure 4.9: Trusted Prior Contacts [Krebs, 2002, pg. 46]

therefore applying RBAP to this data is in the context of centrality rather than status or power.

An initial look at the rank orderings based upon RBAP scores and varying levels of α are provided in Figure 4.10; higher RBAP scores result in higher rankings which for this network range from low (1) to high (19). As observed with the Katz data, determining the most central individuals according to the RBAP measure is predicated upon the amount of attenuation assumed. Mohammed Atta (actor 5), the purported ring leader, is initially tied with seven other individuals, all having a degree of 3, for rank 9. However, as α is increased to 1, meaning less attenuation with longer paths, Atta's rank goes down substantially. Crossovers such as these

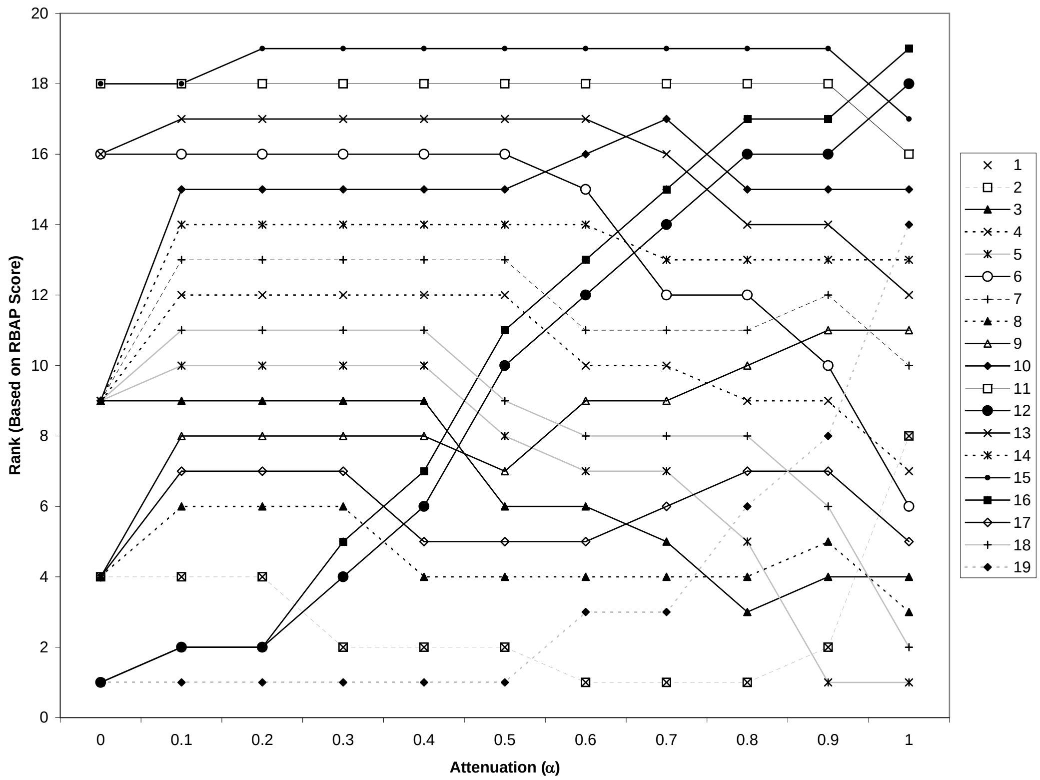


Figure 4.10: RBAP and Hijacker Network

may reveal individuals that are strategically connected from a local perspective, but effectively cut off from the remainder of the network from a global perspective. Such an individual could serve as a cell or team coordinator practicing good operational security techniques, or may be a specialist kept at a distance to the main group. If they correspond to a critical skill or capability, they may be the trigger mechanism for an impending operation.

Other interesting results from Figure 4.10 are those actors who remain low (or high) regardless of the level of α as well as those who start low at $\alpha = 0$ and move up with increasing α. Consider actors 1 Suqami and 2 Wail Alshehri whose RBAP measures tend to stay relatively low over the range of α. These terrorists are not only in the periphery of the network, but they are both connected via actor 3,

Waleed Alshehri, who is also somewhat isolated from the network and whose RBAP measure exhibits similar behavior to that of Atta (decreasing with α).

In contrast, two other apparently isolated actors, A. Alghamdi (12) and S. Alhazmi (16), are connected directly to two of the most central actors (11 and 15) from a betweenness, information centrality, eigenvector, and Katz perspective [verified by Cyram, 2004]. Consequently, despite the low degree of actors (12) and (16), they are connected directly to the core of the network which significantly improves their corresponding RBAP scores as the impact of attenuation is diminished. Note also that the most central actors, H. Alghamdi (11) and N. Alhazmi (15) not only begin with a high rank (due to their high degree) but maintain their relatively high ranking throughout the range of α.

4.6 RBAP and Layered Networks

When considering layered networks, let $\mathbf{X}_r$ denote the sociomatrix for the relationships $r \in \mathfrak{R}$ of interest to the decision maker or analyst. By this very construction, each $\mathbf{X}_r$ is considered to be a separate network, disconnected from the others due to the different context in which those individuals interact; additionally, some layers may be comprised of multiple components. Recall that RBAP should be applied only to connected networks to facilitate interpretation of the results. Nonetheless, two different approaches to the use of RBAP on layered networks are conceived.

The first, is to apply RBAP to each, presumably connected (i.e, a single component) layer independently. Actors demonstrating high centrality, power, or status in each layer could be candidates for further scrutiny. Actors demonstrating these characteristics in more than one layer would therefore be of even more interest. Direct comparison of RBAP results between two or more layers, specifically the values, should only be accomplished if each of those layers is connected and has the same

number of actors, N. This is due the the connectedness requirement as well as the limiting path length of $(N-1)$.

The second approach involves an aggregation of network layers prior to implementing the RBAP measure. Since the input required is merely a dichotomous representation of the network, information characterizing the contexts and potentially different contributions to the strength of interpersonal relationships is regrettably lost, a criticism of all other social network analysis measures using similar inputs within the literature. Nonetheless, the underlying motivation of serving as an initial screening methodology, it is assumed that this is a reasonable tradeoff in some situations. Methods such as the derivation of multiplex-indices may be used to aggregate layered networks into a single, dichotomous representation of the network, capturing the overall effects of multiple relationship contexts [Wasserman and Faust, 1994, pg. 219]. As an example, a simple boolean summation, denoted $x^{\oplus}$, of all relationships, $r \in \mathfrak{R}$, for each possible relationship, $x(i,j) \in \mathbf{X}$ is mathematically shown in Equation 4.5.

$$x\left(i,j\right)^{\oplus} = \begin{cases} 1 \quad if \quad \sum_r x\left(i,j\right)_r \geqslant 1 \\ 0 \quad if \quad \sum_r x\left(i,j\right)_r = 0 \end{cases} \tag{4.5}$$

Borrowing from Wasserman and Faust [1994], a conditional boolean summation would require a threshold for a relationship to exist between two individuals i and j in m or more contexts; this approach is mathematically defined by Equation

$$x\left(i,j\right)^{\oplus} = \begin{cases} 1 \quad if \quad \sum_r x\left(i,j\right)_r \geqslant m \\ 0 \quad if \quad \sum_r x\left(i,j\right)_r = 0 \end{cases} \tag{4.6}$$

This second technique essentially acts as a filter, including only those interpersonal ties that may be significantly stronger than others due to their inherent multiplexity.

Suppose that if certain patterns of interaction, represented by a subset or combination of subsets of relationships, were sufficient (or required) to indicate an interpersonal tie that should be included in further analysis. These patterns could

also be used to filter the necessary relationships for input into RBAP. For example, assume that a relationship pattern of interest, $\{\wp\}$, is defined as a subset of the individuals relations, $r \in \mathfrak{R}$, such that $\{\wp\} \subseteq \{\mathfrak{R}\}$. This subset $\{\wp\}$ could be comprised of one or more of the relations, r. In addition, there may be one or more patterns of interest, $p = 1, \ldots, P$. Therefore, a third method to aggregate multiple contexts into a single sociomatrix for input into RBAP is defined:

$$x\left(i,j\right)^{\oplus} = \begin{cases} 1 \quad if \quad x\left(i,j\right)_r = 1 \; \forall \text{ r} \in \wp_p, \forall \; p \in P \\ 0 \qquad\qquad\qquad \text{otherwise.} \end{cases} \tag{4.7}$$

RBAP may be compared to the results of other measures through the use of nonparametric statistics such as Spearman's coefficient of rank correlation [Lind et al., 2002, pg. 605]. With n denoting the number of paired observations and d denoting the difference between the ranks for each pair, the coefficient is defined as

$$r_s = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)}. \tag{4.8}$$

To test the statistical significance of r_s, assuming the network size is comprised of 10 or more individuals, the test statistic, t, is calculated by

$$t = r_s\sqrt{\frac{n-2}{1-r_s^2}} \tag{4.9}$$

and then compared to the Student-t value with a desired level of significance and $(n-2)$ degrees of freedom [Lind et al., 2002, pg. 607].

The Wilcoxon signed-rank test seeks to find statistically significant changes in the rank orderings of a given set of individuals. This nonparametric method is "based on the differences in dependent samples, where the normality assumption is not required" [Lind et al., 2002, pg, 591]. Since the null hypothesis is defined as *no difference in the rankings*, a rejection of this hypothesis could be used in at

least two different ways. The first would facilitate arguments about the specified level of α desired as input for the RBAP measure. If uncertainty or argument regarding a specific value for α is prevalent, then this statistical technique may assist in ascertaining whether the subsequent changes in individual rankings due to a minor change in α is significant. If the hypothesis is rejected, then argument and further analysis of α is warranted due to its impact upon the statistically significant differences in actor rankings. If the method fails to reject the hypothesis, there is insufficient evidence to conclude that there is a statistically significant difference between the two rank orderings. However, decision makers must assess the tradeoffs between further data collection or continuing with an agreed-upon level of α in the interim.

The second potential application of this method would facilitate change detection. Suppose that the α level was agreed upon and constant data collection regarding the relations among a set of actors was being collected. Increased or decreased activity, and the concomitant links created or removed among the members, could result in changing positions (RBAP values) among the set of actors of interest. Changes in roles, responsibilities, communication activity and patterns, and so forth, may then be captured by these varying RBAP values. Statistical comparisons between previous and new rankings based upon the RBAP values offer a means to detect a statistically significant change in actor centrality, ceteris paribus. Detection of this event would then facilitate when further analysis and/or intelligence resources are required.

4.7 Summary

The measure presented in this chapter shares some aspects of other walk and path-based approaches to gaining insight into an actor's potential for influence or power based upon their position within a given network. However, RBAP provides more analytic freedom regarding the common assumption of attenuation as a function

of distance between individuals. In addition, the small-world property often inherent to social networks provides a degree of computational efficiency to the measure. Consequently, assuming that the network of interest is reasonably sized (e.g., 3000 actors or fewer) this measure should be responsive to changing information.

The intended purpose for RBAP is to facilitate the investigation of adversarial non-cooperative networks, particularly if the network consists of large number of actors. Actors of interest may be identified by consistently high or low RBAP scores as well as those that improve or decrease significantly with a corresponding change in α. Those individuals that are identified through this process can then be subject to increased intelligence scrutiny, either to improve the accuracy of the network data, or to set the stage to affect the organization for political purposes.

Such political endeavors often involve persuading an organization to change position on a given issue, to modify the inherent approach used to achieve their goals, or to even disband entirely. Given that an adequate amount of information regarding the individuals and their associated relationships has been obtained, courses of action to achieve these political endeavors could include persuading the entire organization from within. For example, assuming the clandestine network is adversarial, one must first determine those individuals that are accessible. Among this set, those with higher RBAP scores, and who are consequently more effective at reaching or influencing others, would make attractive participants of collusion.

Although α has been specified as a scalar to this point, a possible extension of this measure could incorporate a matrix of individual-specific attenuation factors. Therefore each individual i would be assigned an attenuation factor, α_i. The scalar α in Equation 4.2 would simply be replaced by the diagonal matrix $\mathbf{A}$, where $\mathbf{A}(i,i) = \alpha_i$, zero otherwise. A possible means to estimate these values could be derived via a decision analytic model using the five bases for power—attraction, expert, reward, coercive, and legitimate—specified by [French, 1956, pg. 183-5] or individual characteristics such as charisma, appearance, and so forth. This data

could then be used to gain insight into the effort required to discredit (or support) a specific individual, thereby diminishing (or increasing) their relative power or influence within the network. Holding all other individuals' attenuation factors constant, sensitivity analysis of the attenuation factor of the individual of interest would yield the concomitant change in power structure based upon the RBAP scores.

From a counter-terrorism perspective, the RBAP measure offers another means to gain insight into adversarial, clandestine networks such as Al Qaeda, Ansar al Islam, and the many others that threaten peace. Due to the secretive nature inherent to these organizations, methods that provide useful information despite limited or uncertain data are of interest. From a social networks perspective, this measure is not intended to be a direct competitor to the numerous, classical measures in existence, but a complement to enhance the structural study of network data. It should be noted that the RBAP measure may also be applied to physical networks–layers of interrelated infrastructure networks, for example. Identifying well-connected, critical nodes in such networks is a key operational consideration in today's security environment.

The MATLAB code developed to perform the RBAP measure, which accommodates either the scalar α or the matrix $\mathbf{A}$, and a sensitivity analysis procedure for α is provided in Appendices A and B, respectively.

Suppose now that sufficient information has been obtained on the network of interest, accurately identifying the individuals of interest and the interpersonal relationships among them. A next step could include the evaluation of targeting options. Although targeting in the military traditionally refers to a process that ultimately results in the physical disruption, or more often destruction, the focus instead is upon the application of influence operations upon individuals within a target network. Such operations seek to affect an adversary's decisions through various means of influence. An analysis methodology to ascertain which subset of

actors should be targeted in order to influence the overall network is discussed in Chapter V.

V. Target Development

5.1 Chapter Overview

Consider the objective of indirectly influencing an entire network of individuals by directly influencing a subset of its actors. One possible means of achieving this objective is the application of the second of the two key player problems. Recall from Section 2.2.2, the key player problem (KPP) involves finding certain *key* individuals within a given social network. How key these players are is dependent upon one of two objectives. The first objective, denoted KPP-1, determines which set of individuals that, once removed, would cause the most damage to network. In this context, Borgatti [2006] defines *damage* as either increasing the number of components within the graph or, if that is not possible, significantly increasing the distance between all pairs of nodes. The second objective, denoted KPP-2, determines which set of individuals that, if successfully convinced to do so, can reach out and influence the majority or, if possible, the entirety, of the other members within the network.

Clearly, both of these problems have military applications. However, the focus of this research is KPP-2 and its use in efficiently propagating influence via direct or indirect contact among individuals within a target network, hence the key player influence problem. From Definition 8, KPP-2 seeks a *kp*-set of order K that is maximally connected to all other nodes [Borgatti, 2006].

As an example, KPP-2 could be used to select individuals that serve as the optimal targets of influence operations [cf. Borgatti, 2003a, 2006]. In this context, optimality is based upon minimizing both the cardinality of the set as well as the distance the influence emanating from an individual must traverse to reach the assigned contact. An example of propagated influence could include the execution of a psychological operation (PSYOP) or a more specific application of PSYOP, an influence operation or campaign. PSYOP is formally defined as

> "...planned operations to convey selected information and indicators to foreign audiences to influence the emotions, motives, objective reasoning, and ultimately the behavior of foreign governments, organizations, groups, and individuals" [DOD, 2003, pg. ix].

KPP-2 offers a means to select such a set of individuals in an efficient manner, essentially supporting the *target development* and *target value analysis* processes. The target development process, from the perspective of Joint Doctrine, "examines potential adversary military, political, or economic target systems to identify subcomponents or elements and interrelationships" [DOD, 2002, pg. C-6]. Target value analysis "establishes criticality of a target or target system in order to select candidate aimpoints that should be attacked to achieve desired effects and accomplishes the defined objectives" [DOD, 2002, pg. C-6]. After a brief description of the underlying motivations behind the KPP measures and the current solution approaches proposed by Borgatti [2006], several mathematical programming extensions and examples are offered and discussed in detail.

5.2 Background

Borgatti [2006] and Everett and Borgatti [1999] note that the preponderance of social network measures focus on the characterization of individual actors and their role or position within the network. Consequently, when the role or position of a group of actors is sought, the traditionally actor-specific measures fail to provide analysts with the proper insights for group-specific results. In particular, the traditional social network analysis measures cannot account for the effects of redundant and structurally equivalent actors in a group, a subset of the social network. Therefore, the key player problem construct was devised.

Both maximal disruption and maximal connection are topics of military interest. For example, application of KPP-1 could result in the disruption of terrorists' social networks, thereby (potentially) impeding their ability to coordinate, plan, and

execute future operations. It should be noted, however, that the fracturing of some networks into components may actually reduce decision and reaction times if the group members are no longer required or do not feel compelled to coordinate with central authority. It is important to understand the groups' operating doctrine prior to pursuing any destructive efforts. Nonetheless, what is of primary interest in this research is KPP-2, the application of which could include identifying a group of actors, a subset of a larger target network, to serve as a conduit to the remaining members.

Such an approach would seek to influence, via the spread of a message or information operations product or campaign, the entirety of network membership in an efficient manner. Of course, KPP-2 could also be viewed in an abstract fashion with regard to the type of network under study by seeking to efficiently influence networks beyond those strictly limited to interpersonal interaction. For example, demographic strata within a population, cities within a country, countries within a region, or components within physical infrastructure, could all be represented by nodes within a network within which information or influence flows.

Similar approaches have been studied in the area of *viral marketing*, essentially an electronic or Internet-based version of "word of mouth" advertising techniques. The underlying premise of this approach is that

> "...by initially targeting a few influential members of the network–say, giving them free samples of the product–we can trigger a cascade of influence by which friends will recommend the product to other friends, and many individuals will ultimately try [the product]" [Kempe et al., 2003, pg. 137].

Determining which individuals serve as ideal initial targets in the context of marketing has been studied by Domingos and Richardson [2001], among others. For example, comparing the marketing costs incurred to reach a given individual to "the expected profit from the sales to other customers she may influence to buy, the customers those may influence, and so on recursively" would facilitate the cost-effective

selection of key individuals [Domingos and Richardson, 2001, pg. 57]. Extending similar problem aspects within the setting of an influence campaign, such as cost to access, surveil, or turn a specific individual to do another's bidding, offers an opportunity to further develop the key player problem paradigm.

Although these efforts share common themes, note that the specific definition of *efficiency* may vary between applications and decision makers. Borgatti's approach and concomitant assumptions, discussed next, serve as the initial basis for comparison against the proposed mathematical programming formulations.

5.3 Heuristic Approach and Objective Function

The current options available within the published KPP software attempt to satisfy KPP-2 from three approaches. The first two are derived from different approaches to the number of nodes reached criterion. The third option seeks to minimize the measure Borgatti refers to as reciprocal distance reach. For the *number of nodes reached* criterion, the user specifies the maximum number of degrees of separation allowed between a key player and its assigned member as well as either a specific *kp*-set size or a maximum allowable size for the *kp*-set. The *specific size setting* then seeks to reach as many of the other actors as possible, given the *kp*-set and reach constraints. The *maximum allowable size* option selects members until either the entire network is reached, given the conditions specified, or the maximum allowable group size has been allocated. For the *reciprocal distance reach* approach, the user specifies a *kp*-set size, and the software seeks to optimize the measure of efficiency described in Equation 5.1 via a greedy heuristic approach [Borgatti, 2006].

$$\texttt{Max}\ D_R = \frac{1}{N}\sum_j \frac{1}{d_{Kj}} \tag{5.1}$$

For each actor k in a given *kp*-set (K), the distance from all actors j to the closest key player is denoted d_{Kj}. Distances between a key player and itself are

assumed to be 1; the measure is normalized by dividing the summation by the total number of actors in the network (n) [Borgatti, 2006]. Regarding the maximum distance between key players and other actors, Borgatti also recommends limiting this distance to within two degrees of separation. This is not an unreasonable assumption as it has been observed in a variety of communications literature that the farther that information must travel, the more likely that the information content can be misinterpreted, subject to errors or transmission failure, or a combination thereof [cf. Katz, 1953; Stephenson and Zelen, 1989]. This restrictive assumption also limits the possible solutions available; however, it may be relaxed up to a reach of $(n-1)$ for all of the models presented in this chapter. For a more complete discussion on the heuristic methodology, the reader is referred to Borgatti [2006].

Since Equation 5.1 normalized by N, Borgatti contends that $D_R \in [0,1]$. However, while the distance assumption of unity between key players and themselves along with the assumption that $d_{Kj} = \infty$ for any actor unreachable by any key player suggest that $\lim_{d_{Kj}\to\infty} \frac{1}{d_{Kj}} = 0$ for unreachable actors, the appropriate range is actually $D_R \in [(K/N), 1]$. Nonetheless, the overall objective is to maximize this function.

For the *number of nodes reached* criterion, it is shown that equivalent integer programs, specifically the minimum covering and fractional covering problems, can be formulated and solved to optimality in reasonable amounts of time. For the *reciprocal distance reach* criterion, variations of the p-median and the related facility location problem can be applied to address various aspects of this problem, as well as optimize the objective given in Equation 5.1. However, since it is assumed that a decision maker seeks to influence a particular network, potential comparisons between different networks, and therefore the normalization of Equation 5.1 by N, is no longer necessary. Additionally, D_R meets the (applicable) assumptions of proportionality and additivity required of linear programming [Hillier and Lieberman,

1995, pg. 38-43]. Applying the information contained within the adjacency matrix $\mathbf{X}$ towards a mathematical programming approach is discussed next.

5.4 Mathematical Formulations

Recall that, given a network of interest, the cells $x(i, j) \in \mathbf{X}$ are defined as $x(i, j) = 1$ if there exists an arc or relationship between actors i and j, 0 otherwise, constituting a simple adjacency matrix in operations research terms. Consequently, each nonzero entry in $\mathbf{X}$ implies that i can reach j in one step. Similarly, the cells of the square of the adjacency matrix, denoted $\mathbf{X}^2$, indicate the number of directed edge sequences of length two from i to j. In this particular case, these sequences are equivalent to paths. Therefore, for all $i \neq j$, $x(i, j) > 0 \in \mathbf{X}^2$ implies that there exists a path from i through some intermediary node or actor, and then on to j [Deo, 1974].

The transpose of the matrices described above (denoted $\mathbf{X}'$) does not change the graph, but it does change the interpretation of the information contained within $\mathbf{X}$. Given the transpose, the columns (as opposed to the rows) of $\mathbf{X}'$ correspond to whom the individual can reach in one step (e.g., adjacent actors). Let $\mathbf{R}_1 = \mathbf{X}' + \mathbf{I}$ denote the modified one-step reachability matrix. Note that the ith column of $\mathbf{R}_1$ captures the nodes that are adjacent to actor i, including itself. Additionally, taking the transpose is only necessary if the graph is asymmetric, which implies that this approach can be readily extended to directed graphs–an improvement over the current methods proposed by Borgatti [2003b].

This approach may also be applied when a reach of more than one step away is allowed. For example, let $\mathbf{R}_2 = \delta[(\mathbf{X}')^2 + \mathbf{R}_1]$ denote the two-step or less reachability matrix, where in general $\delta(a) = 1$ if $a > 0$, zero otherwise. Consequently, the ith column of $\mathbf{R}_2$ captures the nodes that are at most two-steps away from actor i, including itself. This could be extended for reachability via longer path lengths using

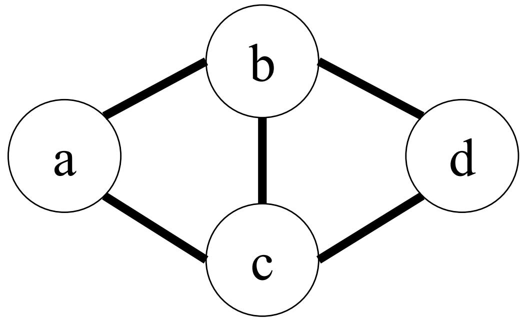


Figure 5.1: Notional Network and Coverings

the technique described in Wasserman and Faust [1994, pg. 160]. These matrices can then be used to form the constraint matrices of the various mathematical programs (MP) proposed in this study which address the two variants and extensions of KPP-2 investigated by Borgatti [2006]: the number of nodes reached and the reciprocal distance reach. The multitude of extensions via mathematical programming techniques, several of which are presented here, enable a more comprehensive analysis capability.

5.4.1 Number of Nodes Reached

Borgatti likens the *number of nodes reached* approach to the classical graph theoretic vertex cover problem, which is "a subset of vertices that includes at least one vertex incident on every edge of [the graph]" [Deo, 1974, pg. 193] However, KPP-2 is actually more closely related to the edge covering problem, which is defined as a set of edges to which every vertex in the graph is incident to at least one edge [Deo, 1974, pg. 182]. Consequently, the edges induced from the *kp*-set and their assignments to the remaining members generates the edge cover. To illustrate these subtle differences, consider the graph in Figure 5.1.

A vertex cover for this graph is the set $\{b, c\}$; at least one of these two nodes is incident to all edges within the graph. However, there are two optimal KPP-2 solutions, from both the number of nodes reached and the reciprocal distance

perspectives, comprised of either $\{b\}$ or $\{c\}$. Node b, for example, can reach all other actors within one step. Assuming that b is selected as the key player, note that edges (a, c) and (c, d) are not incident to b, therefore the *kp*-set solution is not a vertex cover. However, assume again that node b is selected as the *kp*-set solution; the edges used to assign node b to the remaining individuals in the network $\{(b, a), (b, c), (b, d)\}$ form an edge cover, since all vertices are adjacent to at least one of these edges.

Borgatti also relates dominating sets to KPP-2, but suggests that this method fails due to ...

> "The focus of graph-theoretic research has been on finding the smallest cover or dominating set that achieves the goal (reaching all nodes) perfectly. [The KPP-2] problem is the reverse: finding a set of fixed size that achieves the goal as well as possible. In addition, we would prefer to measure the extent to which a set reaches all nodes, so that we can evaluate our success" [Borgatti, 2006].

Viewing the graph theoretic approach of dominating sets from a different perspective, however, may still yield valuable insight into KPP-2. First, recall that a *dominating set* is "a set of vertices that dominates every vertex [in a graph] in the following sense: Either [a vertex] is included in the dominating set or is adjacent to one or more vertices included in the dominating set" [Deo, 1974, pg. 172]. Further, a *minimal dominating set* is "a dominating set from which no vertex can be removed without destroying its dominance property" [Deo, 1974, pg. 172]. Note that minimal dominating sets can be of varying sizes, the smallest of which is generally referred to the *minimum dominating set* whose cardinality represents the domination number (or domatic number) of the graph [Deo, 1974, pg. 173]. An extension of the dominating set, and an accompanying distributed algorithm to solve it, is the k-dominating set presented in Penso and Barbosa [2004]. Given an integer $k \in [1, (N-1)]$, all nodes are either in the dominating set, or at most a distance of k steps away from the nearest node within the dominating set [Penso and Barbosa, 2004, pg. 243]. Con-

sequently, the k-dominating set is another means of describing the KPP-2 problem, where a reach distance of more than one step is allowed.

Given that the KPP-2 problem seeks to maximally connect a kp-set to the remaining network with as few key players as possible, the minimum dominating and the minimum k-dominating sets are appropriate methodologies to analyze KPP-2. Next, relations between the dominating set and the minimum set covering problem (MCP) formulations are reviewed. Additionally, Borgatti's concerns regarding a 'perfect' or 'complete' cover can be addressed by modifying the MCP to represent a fractional covering problem (FCP), respectively.

The minimum covering problem seeks to select a minimum number of objects that can cover a set of interest. Examples include the minimum number of personnel required to meet shift requirements or the determination of facility locations in order to meet certain demands (e.g., fire stations and minimum response times for various sections of a city) [Nemhauser and Wolsey, 1999, pg. 6-7]. In this case, the decision is to designate a portion of individuals within the network, the kp-set, to 'cover' themselves and as many of the other network members by having the ability to communicate with them either directly or through at most one intermediary. The general formulation for the minimum cover is

$$\texttt{(MCP)} \qquad \texttt{Min} \ \left\{ z = \mathbf{cx} : \mathbf{Ax} \geq \mathbf{1}, x_i \in \{0,1\}, \forall i \right\}. \tag{5.2}$$

When the constraint matrix $\mathbf{A}$ in Equation 5.2 is simply replaced with $\mathbf{R_1}$, the MCP solution provides the minimum dominating set for the graph [Christofides, 1975]. Note that this approach also lends itself to k-dominating sets via the use of $\mathbf{R}_m$ where $m > 1$, as well as directed graphs due to the incorporation of $\mathbf{X}'$ within $\mathbf{R}_m$.

For a majority of the formulations presented in this chapter, the main decision variable is x_{i_m}, which equals 1 if actor i is chosen as a key player and must reach its

assigned members within m steps and 0 otherwise. For notational convenience, let the vector $\mathbf{x_m}$ represent x_{i_m} for all $i = 1 \ldots N$. Taking the number of nodes reached approach, determining the *kp*-set that is no more than one step away from all other actors is accomplished by the mathematical program in Equations 5.3 and 5.4. This approach is denoted (NR1), the *kp*-set that satisfies the *number reached* approach, for all actors, within one step. Note that the cost associated with selecting any given actor is assumed to be 1, so that this approach seeks to minimize the cardinality of the *kp*-set. Consequently, this approach also finds the minimum dominating set.

Since all of the formulations presented are essentially variants of the covering problem, the cost coefficients are not required to be 1. As an example, the (DNR1) formulation, shown later, distributes the workload among key players as evenly as possible by taking into account a member's adjacent nodes (see Equation 5.18). Therefore, when trying to choose members as key players, all formulations presented in this chapter are amenable to extensions that incorporate various costs that may be associated with designating a key player or not being able to cover a particular individual.

For the network shown in Figure 5.2, there are multiple optimal solutions to the (NR1) problem, namely, $\{a, c, e\}$, $\{a, c, f\}$, $\{a, d, e\}$, $\{a, d, f\}$, $\{b, c, e\}$, $\{b, c, f\}$, $\{b, d, e\}$, and $\{b, d, f\}$. Although a majority of mathematical programming software suites typically only output a single solution (among multiple optima), there are straightforward techniques to identify the existence of and to enumerate multiple optima. Additionally, a method for enumerating all minimal dominating sets using Boolean arithmetic can be found in [Deo, 1974, pg. 173]. Since a key player is expected to be part of the influence operation, (NR1) will always be feasible despite

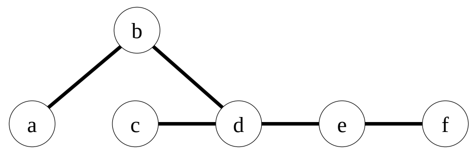


Figure 5.2: Notional Network

the existence of isolated nodes or analysis of a graph with multiple components.

$$\texttt{(NR1)} \qquad \texttt{Min} \sum_{i=1}^{N} x_{i_1} \tag{5.3}$$

$$\texttt{Subject To } \mathbf{R}_1\mathbf{x}_1 \geq \mathbf{1} \tag{5.4}$$

$$x_{i_1} \in \mathbb{B} \quad \forall \; i$$

If the requirement that the key players must be adjacent to the remainder of the network can be relaxed, then a covering problem representation of the minimum k-dominating set can be applied. Letting the vector $\mathbf{x_2}$ represent x_{i_2} for all $i = 1 \ldots N$ and incorporating the matrix $\mathbf{R}_2$ described earlier, this approach permits indirect influence to occur between key players and other network members as long as all members are two steps or less from a key player. This mathematical program, denoted (NR2), also exhibits the same goal of minimizing the cardinality of the *kp*-set and is shown in Equations 5.5 and 5.6. Applying this approach again to the network shown in Figure 5.2, the (only) optimal *kp*-set is comprised of actor $\{d\}$. The objective functions of (NR1) and (NR2) specify which actors should comprise the *kp*-set and the extent of reach that is required of them in order to maximally connect to the entire network. Applying techniques used to evaluate reachability in dichotomous social networks, any distance $m \in [1, (N-1)]$ could be used [Wasserman

and Faust, 1994, pg. 160].

$$\texttt{(NR2)} \qquad \texttt{Min} \sum_{i=1}^{N} \sum_{m=1}^{2} x_{i_m} \tag{5.5}$$

$$\texttt{Subject To} \quad [\mathbf{R}_1 \,|\mathbf{R}_2] \, [\mathbf{x}_1 \,|\mathbf{x}_2] \;\geq\; \mathbf{1} \tag{5.6}$$

$$x_{i_m} \;\in\; \mathbb{B} \quad \forall \; i, m$$

As suggested previously, a fractional covering problem approach may be used to account for situations when the entire network cannot be reached within m steps or less if the size of the *kp*-set is limited (e.g., constrained resources limit the number of individuals that may be coopted). Borrowing from the work of Gandhi et al. [2004], let $y_i = 1$ if node i is covered, 0 otherwise. The resulting fractional covering problem approach to (NR1), denoted (FNR1), Equations 5.7 through 5.9, where U is the maximum number of actors that may be missed or not covered by at least one of the key players.

$$\texttt{(FNR1)} \qquad \texttt{Min} \sum_{i=1}^{N} x_{i_1} \tag{5.7}$$

$$\texttt{Subject To} \quad \mathbf{R}_1 \mathbf{x}_1 + \mathbf{y} \;\geq\; \mathbf{1} \tag{5.8}$$

$$\sum_{i=1}^{N} y_i \;\leq\; U \tag{5.9}$$

$$x_{i_1}, y_i \;\in\; \mathbb{B} \quad \forall \; i$$

The primary difference between (NR1) and (FNR1) is the option to not cover a given number of nodes U in the event that the size of the *kp*-set does not permit full access to all actors within the network. Of course, this approach can be applied to the *number of nodes reached* problem where more than one degree of separation

is permitted. The formulation for no more than two degrees of separation is:

$$
\begin{aligned}
\texttt{(FNR2)} \qquad\qquad \texttt{Min} \; \sum_{i=1}^{N}\sum_{m=1}^{2} x_{i_m} & & & \qquad (5.10)\\
\texttt{Subject To} \;\; [\mathbf{R}_1\,|\mathbf{R}_2]\,[\mathbf{x}_1\,|\mathbf{x}_2] + \mathbf{y} &\geq \mathbf{1} & & \qquad (5.11)\\
\sum_{i=1}^{N} y_i &\leq U & & \qquad (5.12)\\
x_{i_m}, y_i &\in \mathbb{B} & \forall\; i, m. &
\end{aligned}
$$

The objective functions of (FNR1) and (FNR2) specify the smallest set of actors that should comprise the *kp*-set and the extent of reach that is required of them in order to maximally connect a portion of the network. In this fractional case, the *kp*-set can miss no more than $\frac{U}{N}\%$ of the membership. Note that when $U = 0$, both (FNR1) and (FNR2) reduce to (NR1) and (NR2), respectively. Further insights may be gained from the optimization results, such as the search for multiple optima, discussed earlier. Costs associated with designating an individual as a key player, as well as those incurred due to missing an actor may also be included into either (FNR1) or (FNR2). For example, using the (FNR1) formulation, suppose the cost to co-opt an actor i is denoted c_i and the cost to miss an actor i is denoted m_i. Then the objective function, Equation 5.7, would have the form

$$\texttt{Min} \; \sum_{i=1}^{N} (c_i x_{i_1} + m_i y_i). \qquad (5.13)$$

Another formulation that leverages the fractional set covering problem, and also closely mirrors the key player approach suggested by Borgatti [2003a], maximizes the amount of network members covered via a *kp*-set of a specified size, K. Using the 1-step assumption as an example, this problem, denoted (FNRK1), is shown in Equations 5.14 through 5.16. The objective function minimizes the number of actors missed or not covered given that the size of the *kp*-set restricted to exactly

K members. It is assumed that K is less than the cardinality of the minimum dominating set; therefore, the decision maker seeks to take as much advantage as possible via limited resources. If K is greater than or equal to the cardinality of the minimum dominating set, then the restriction imposed by specifying K no longer applies, and (FNRK1) essentially solves (NR1).

$$
\begin{aligned}
\texttt{(FNRK1)} \qquad\qquad & \texttt{Min} \sum_{i=1}^{N} y_i && && (5.14)\\
\texttt{Subject To } \mathbf{R}_1\mathbf{x}_1 + \mathbf{y} & \geq \mathbf{1} && && (5.15)\\
\sum_{i=1}^{N} x_{i_1} & = K && && (5.16)\\
x_{i_1}, y_i & \in \mathbb{B} && \forall \; i &&
\end{aligned}
$$

With the exception of (FNRK1), the optimization focus thus far has primarily been upon minimizing the cardinality of the *kp*-sets while covering as many of the other network members as possible. Consequently, such solutions tend to heavily rely upon a few, oftentimes well-connected, individuals to influence the remainder of the network. There are potential drawbacks to these types of solutions; for example, they do not take into account actor characteristics (e.g., known roles within a network, shared or familial ties, and so forth) and the fact that a mathematical solution may not always coincide with a practical and implementable one (e.g., the workload expected of key players in order to disseminate influence or educate others may be unreasonable). However, the flexibility of the mathematical programming techniques offers ways to avoid these pitfalls.

For example, the models presented can incorporate limitations on the *kp*-sets and/or actors available for selection. Assume that a decision maker wanted to limit the number of individuals with redundant skills [Borgatti, 2003a, pg. 130]. Let individuals 1, 2 and 3 be actors with redundant skills. The decision maker wants to include at most one of these individuals in the *kp*-set and is relying upon the 1-step

assumption. This restriction can be included as an additional constraint, within any of the formulations presented thus far, in the form of $x_{1_1} + x_{2_1} + x_{3_1} \leq 1$. Such a constraint is a slight modification of a type 1 special ordered constraint, often referred to as a multiple choice constraint because only one of the decision variables may equal 1 for any given solution [Martin, 1999, pg. 329]. The general form of allowing at most a individuals out of a specified group g, where the reach conditions m are specified by the user or situation, is given by Equation 5.17. Such a constraint is also related to those used in knapsack formulations. This type of constraint can be applied to as many groups or conditions as necessary, each scenario represented by an additional constraint. In fact, implementation of this type of constraint permits the identification of multiple optima for the formulations presented in this chapter.

$$\sum_{i=1}^{g} x_{i_m} \leq a \tag{5.17}$$

As another example of the flexibility inherent within the MP approach, assume that the decision maker wanted to ensure that two specific individuals i and j, with $m = 1$, were either both included or both excluded within the *kp*-set. This requirement could be modeled as $x_{i_1} = x_{j_1}$. Ensuring that a specific actor i is included in or excluded from the *kp*-set merely requires the additional constraint $x_{i_1} = 1$ or $x_{i_1} = 0$, respectively. Requiring i or j would be modeled as $x_{i_1} = 1 - x_{j_1}$. This could also be used to intentionally forego subjecting an actor or actors to an influence campaign, particularly if they are the ultimate target of a coup or insurrection by the remaining group members. Along these lines, and in the case of fractional coverings, the analysis of the solution and decision variables also permits the investigation of individuals that may consistently be missed despite multiple optima.

Clearly, the options are limited only by the analyst's requirements and ingenuity. To deal with key players that are overburdened simply due to their connectivity,

cost coefficients (implicitly assumed to be 1 to this point) may be included in the objective function in order to 'spread out' the work required of the *kp*-set membership. A means to accomplish this is discussed next.

5.4.1.1 Leveraging Key Player Costs

The covering approaches (NR1) and (NR2) could easily incorporate individual-specific costs, represented by c_{i_m}, which denotes the cost associated with selecting actor i as a member of the *kp*-set and assigning them to all possible actors within m steps or less. Such costs may account for the time, effort, or resources required to successfully recruit the actor into the *kp*-set. Alternatively, the cost could represent a function of the demand placed upon the individual once they are included in the *kp*-set (e.g., the key player is required to distribute informational pamphlets to all of its assigned members).

Continuing with the 1-step assumption, let $c_{i_1} = 1/d(i)_{out}$, where $d(i)_{out}$ is the out-degree of actor i. With the objective of minimizing such costs, this may initially appear as self-defeating, since greater out-degrees, and therefore the more workload potentially imposed, result in smaller cost coefficients. However, introducing a constraint specifying the size of the *kp*-set results in more evenly distributed expectations upon the set members.

For example, assuming that a decision maker wants to cover all N actors in the network within 1-step, suppose a *kp*-set of size $K = 2$ were desired. Let a and b represent the out-degrees of each actor of an arbitrary *kp*-set solution. An additional constraint is added to the (NR1) formulation such that $\sum_i x_{i_1} = K$. This modified formulation providing a *kp*-set of order K with an evenly distributed workload to reach all other actors within one step is denoted (DNR1) and is shown in Equations 5.18 through 5.20. Note that formulations amenable to this approach must require K to be at least as great as the domatic number of the graph being analyzed.

The smallest possible objective function value for this problem occurs when each actor covers an equal (or near equal in the case of an odd number of actors) number of other actors. Assume that a perfect cover is possible such that $a + b = N$ (i.e., there is no overlap or redundancy among the key players' assignments), with $a, b \in \mathbb{Z}^+$. The cost associated with any solution of this nature is $\frac{1}{a} + \frac{1}{b}$. If $a = b$, then this reduces to $\frac{2}{a}$. Suppose that $a \neq b$, such that $a = a - \epsilon$ and $b = a + \epsilon$, $\epsilon \in \left[\left(1 - \frac{n}{2}\right), \left(\frac{n}{2} - 1\right)\right]$. Now, $\frac{1}{a} + \frac{1}{b} = \frac{2a}{a^2 - \epsilon^2}$ implies that for all $\epsilon \neq 0$, and therefore any solution with $a \neq b$, is inferior to one with $a = b$. Therefore, this model seeks to distribute, as evenly as possible, the influence or contact 'workload' among the members of the *kp*-set. This is easily extended to *kp*-sets of any arbitrary size K with the minimum objective function bounded below, in general, by $\frac{K^2}{N}$.

$$\texttt{(DNR1)} \qquad \texttt{Min} \sum_{i=1}^{N} c_{i_1} x_{i_1} \tag{5.18}$$

$$\texttt{Subject To } \mathbf{R}_1 \mathbf{x}_1 \geq \mathbf{1} \tag{5.19}$$

$$\sum_{i=1}^{N} x_{i_1} = K \tag{5.20}$$

$$x_{i_1} \in \mathbb{B} \quad \forall \ i$$

In the case of overlapping covers such that $a + b > N$, the smallest possible value for this function occurs similarly when $a = b = N$, which results in an objective function for (DNR1) that is bounded below by K/N. An illustrative example, assuming that $N = 10$ and $K = 2$, with objective functions (z) based upon varying set sizes of a and b, is depicted in Figure 5.3. The intersection of the lines $a = b$, $a + b \geq 10$, and $z = 0.4$, emphasize the distributive effect underlying this approach.

The formulations presented thus far simply determine the key players required to influence or cover by contact either all actors within the network, including themselves, at least once or as many as possible if the *kp*-set size is determined *a priori*. These *number of nodes reached* approaches also take into account the cardinality of

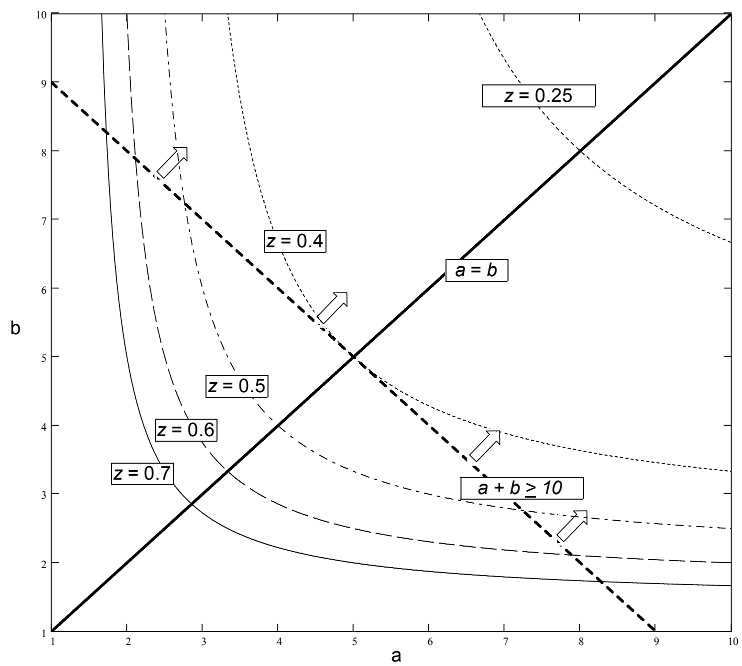


Figure 5.3: Distribution of *kp*-set Effort

the *kp*-set, as well as the the influence workload amongst its members. The attention is not turned to the 'reciprocal distance reach' approach described by Equation 5.1.

5.4.2 Reciprocal Distance Reach

Seeking to maximize Equation 5.1, the same result can be achieved by maximizing the non-normalized version, or minimizing the additive inverse of the non-normalized version. Assuming that $d_{Kj} > 0$, this relationship is

$$\texttt{Maximize}\ \frac{1}{N}\sum_j \frac{1}{d_{Kj}} \equiv \texttt{Maximize}\ \sum_j \frac{1}{d_{Kj}} \equiv \texttt{Minimize}\ \sum_j \frac{-1}{d_{Kj}}. \qquad (5.21)$$

If a given actor j cannot be reached by a key player K, the distance is assumed to be $d_{Kj} = \infty$. Therefore, $\lim_{d_{Kj}\to\infty} \frac{1}{d_{Kj}} = 0$ implies that unreachable key player-actor pairs are not considered as an option for the optimization problem. Note that this

function is optimized when all remaining actors are adjacent to one or more key players (smaller denominators are better). Therefore, this implies that, given a specified kp-set of size K, selecting the set that immediately covers a majority of the network is preferred.

The classical p-median formulation, which has been applied to a variety of facility location and related problems, provides a MP equivalent to the reciprocal distance reach methodology. The p-median problem (PMP) essentially minimizes the sum of the distance between kp-set members and their assigned non-kp-set members [Reese, 2005]. In this context, this model seeks to minimize the total distance between the actors and the closest of p key players. Let $k_i = 1$ if node i is designated a member of the kp-set, K, zero otherwise. The shortest path distance (d_{ij}) between all nodes must be calculated. Let $x_{ij} = 1$ if actor i is covered by a key player j, zero otherwise. Borrowing from [Handler and Mirchandani, 1979, pg. 58-60], the p-median formulation that accommodates the key player problem (PMED) is:

$$
\begin{aligned}
\texttt{(PMED) Min} \quad \sum_{i=1}^{N}\sum_{j=1}^{n} -{d_{ij}}^{-1} x_{ij} & & & \qquad (5.22)\\
\texttt{Subject To} \quad \sum_{i=1}^{N} k_i &= K & & \qquad (5.23)\\
\sum_{j=1}^{N} x_{ij} &= 1 & \forall\ i = 1,\ldots,N & \qquad (5.24)\\
x_{ij} &\leq k_j & \forall\ i = 1,\ldots,n; j = 1,\ldots,N & \qquad (5.25)\\
k_i \in \mathbb{Z}, x_{ij} &\in \mathbb{B} & \forall\ i,j. &
\end{aligned}
$$

Continuing with the assumption that distances between individuals are based upon zero-one relations, the distances used in the objective function are simply the path distance between individuals. Given a dichotomous network, a reach matrix would be sufficient to ascertain the values for d_{ij} [Wasserman and Faust, 1994, pg. 159]. Note that an objective function that simply minimizes the total distance

(e.g., $\sum_{i=1}^{N}\sum_{j=1}^{n} d_{ij}x_{ij}$) will identify the same KPP-2 solutions. However, the objective function shown in Equation 5.22 easily permits a direct comparison of objective function values between the mathematical programming approach and the heuristic approach presented by Borgatti [2003b].

Analysis using the (PMED) formulation is not necessarily relegated to interpersonal distances based upon dichotomous relationships. Suppose that, in addition to the existence of a relationship, an estimate of the social distance between individuals i and j, $d_{ij} > 0$ can be obtained [cf. Renfro, 2001]. If such values are available, an all-pairs shortest paths algorithm may be applied to obtain the distance values required for the objective function [Sedgewick, 1984, pg. 492-4]. Alternatively, distances between non-adjacent individuals, assuming valued relations as input, could be developed via the procedure developed by Yang and Knoke [2001].

A weighted version of the the (PMED) problem has been extensively studied, permitting the inclusion of actor-specific data within the objective function [Reese, 2005, pg. 2]. For example, assume that an appropriate method was devised to estimate a value of 'importance' assigned to each actor, denoted $v_i \in \mathbb{R}^+$, such that the larger the value, the more important their inclusion into the *kp*-set. Such assignments could be based upon access, desirability, or ease of coercing that particular individual, are easily incorporated into this model and requiring only a minor adjustment to the objective function of (PMED):

$$\texttt{Min} \ \sum_{i=1}^{N}\sum_{j=1}^{N} -{d_{ij}}^{-1} v_i x_{ij}. \tag{5.26}$$

By design, the (PMED) formulation seeks to cover all actors within the network with a key player. Combining aspects of this model and the fractional models discussed in Section 5.4.1, (PMED) can be modified in a manner similar to that shown in (FNR1) and (FNRK1) so that not all actors have to be reached; this modification

is denoted (FPMED):

$$
\begin{aligned}
\texttt{(FPMED) Min} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} -{d_{ij}}^{-1}x_{ij} & & & \qquad (5.27)\\
\texttt{Subject To} \quad \sum_{i=1}^{N} k_i &= K & & \qquad (5.28)\\
\sum_{j=1}^{N} x_{ij} + y_i &= 1 \;\; \forall \;\; i = 1,\ldots,N & & \qquad (5.29)\\
x_{ij} &\leq k_j \;\; \forall \;\; i = 1,\ldots,N; j = 1,\ldots,N & & \qquad (5.30)\\
\sum_{i=1}^{N} y_i &\leq U & & \qquad (5.31)\\
k_i \in \mathbb{Z}, y_i, x_{ij} &\in \mathbb{B} \;\; \forall \;\; i,j.
\end{aligned}
$$

To this point, both the (PMED) and (FPMED) models assume that all actors can, eventually, reach all others. This may not always be the case, particularly when dealing with directed networks. Accommodating this phenomenon requires a modification of the indices used to ensure the assignments and coverings are possible. Let the set E represent all pairs of actors (i, j) that are reachable to each other within the social network of interest. If it is desired to limit the actors by a given number of m-steps in addition to distance, then the $\mathbf{R}_m$ matrices previously defined could be used to develop the appropriate constraint matrix such that $E \in \mathbf{R}_m$. The generalized form of the (PMED) formulation is given by (PMEDm):

$$
\begin{aligned}
\texttt{(PMEDm) Min} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} -{d_{ij}}^{-1}x_{ij} & & & \qquad (5.32)\\
\texttt{Subject To} \quad \sum_{i=1}^{N} k_i &= K & & \qquad (5.33)\\
\sum_{j \in E} x_{ij} &= 1 \;\; \forall \;\; i = 1,\ldots,N & & \qquad (5.34)\\
x_{ij} &\leq k_j \;\; \forall \;\; (i,j) \in E; j = 1,\ldots,N & & \qquad (5.35)\\
k_i \in \mathbb{Z}, x_{ij} &\in \mathbb{B} \;\; \forall \;\; i,j.
\end{aligned}
$$

A discussion of some advantages of applying mathematical programming techniques to KPP-2 follows. The disadvantages, particularly when compared to the heuristic approach proposed by Borgatti [2003b], lie within the computational requirements of the large-scale integer programs required to model large social networks. Some preliminary assessments of this area are discussed within the context of the example case studies.

5.5 Advantageous Properties of the MP Approach

There are a number of advantages to the mathematical programming approaches offered, key among these the flexibility offered to the analyst. For example, given the formulations presented, KPP-2 analysis is not restricted to symmetric networks. This enables an assessment of organizations where information or influence is directed by design (e.g., a strict chain of command) or where different individuals have different effects upon one another (e.g., a leader may better serve as a key player than an untrustworthy minion).

Through the use of additional constraints, certain target individuals or even groups of individuals can be specified *a priori* for inclusion or exclusion of *kp*-set membership as desired. Additionally, constraints similar to the one in Equation 5.17 can be added to facilitate enumeration of alternate optima, thereby providing planners and decision makers with options regarding potential courses of action. As an example, suppose all optimal solutions of size K were of interest. Let s denote a particular optimal solution. Beginning with the initial optimal solution, let the members of the *kp*-set comprise the set P. Including an additional constraint in the form of

$$\sum_{i \in P} x_{i_m} \leq (k-1), \qquad \texttt{for each } s \tag{5.36}$$

Table 5.1: Goal Formulations [Ignizio, 1982, pg. 377]

Goal Type	Goal Programming Form	Deviation Variables to Be Minimized
$f_i(\mathbf{x}) \leq b_i$	$f_i(\mathbf{x}) + \eta_i - \rho_i = b_i$	ρ_i
$f_i(\mathbf{x}) \geq b_i$	$f_i(\mathbf{x}) + \eta_i - \rho_i = b_i$	η_i
$f_i(\mathbf{x}) = b_i$	$f_i(\mathbf{x}) + \eta_i - \rho_i = b_i$	$\eta_i + \rho_i$

in a cumulative manner for each optimal solution found forces the mathematical program to find any remaining solutions. This procedure is repeated until the problem becomes infeasible, or an appropriate number of solutions has been obtained.

External costs (e.g., operational risks endured to co-opt a particular member, amount of money or goods needed for bribes, the expected time required to successfully co-opt an individual, and so forth) as well as internal costs incurred (e.g., the workload endured by a given *kp*-set member) are all easily included into this analysis by tailoring the objective functions. The (PMED) model, for example, can incorporate fixed costs associated with hiring or co-opting key players, essentially a social network version of the classical facility location problem [Handler and Mirchandani, 1979, pg. 58-59].

Another potential extension involves a goal-programming approach to allow trade-offs between, for example, the size of the *kp*-set and the desired number of individuals reached or influenced. The relaxation of some of these constraints via the use of deviational variables incorporated into the objective function is commonly referred to as goal programming. "An aspiration level is a specific value associated with a desired or acceptable level of achievement of an objective" [Ignizio, 1982, pg. 376]. "An objective in conjunction with an aspiration level is termed a goal" [Ignizio, 1982, pg. 376]. To transform an objective i into a corresponding goal, the deviational variables η_i and ρ_i denote the amount under and over a specified goal, respectively. The various transformations for each type of objective are shown in Table 5.1.

For example, consider the (FNRK1) formulation in Equations 5.14 through 5.16. Suppose that the problem is to evaluate the tradeoffs between kp-set size and the number of actors that may be missed if the kp-set size must be some number less than the network's domatic number. Let the first goal be $f_1(\mathbf{x}) = \sum_{i=1}^{N} x_{i_1} = K$. The second goal is denoted $f_2(\mathbf{y}) = \sum_{i=1}^{N} y_i = U$. Using Table 5.1, the corresponding transformations for goals 1 and 2 are $\sum_{i=1}^{N} x_{i_1} + \eta_1 - \rho_1 = K$ and $\sum_{i=1}^{N} y_i + \eta_2 - \rho_2 = U$, respectively. The resulting goal programming formulation for (FNRK1) is denoted (GPFNRK1):

$$
\begin{aligned}
\texttt{(GPFNRK1)} \qquad & \texttt{Min } \eta_1 + \rho_1 + \eta_2 + \rho_2 && (5.37)\\
& \texttt{Subject To } \mathbf{R}_1\mathbf{x}_1 + \mathbf{y} \geq \mathbf{1} && (5.38)\\
& \sum_{i=1}^{N} x_{i_1} + \eta_1 - \rho_1 = K && (5.39)\\
& \sum_{i=1}^{N} y_i + \eta_2 - \rho_2 = U && (5.40)\\
& x_{i_1}, y_i \in \mathbb{B} \quad \forall\ i\\
& \eta_j, \rho_j \in \mathbb{Z} \quad \forall\ j.
\end{aligned}
$$

Due to the equally weighted deviational variables, the (GPFNRK1) formulation implicitly assumes that meeting either goal is equally desirable. To investigate the impact of changing these weights, $\theta \in [0, 1]$, may be incorporated into the objective function as

$$\texttt{Min } \theta(\eta_1 + \rho_1) + (1 - \theta)(\eta_2 + \rho_2). \tag{5.41}$$

The problem may then be solved for various values for θ, providing solutions (assuming feasibility) that trade off deviations between the two goals.

Upon review of the solution output of a mathematical program, there may be indications of surplus conditions for the model constraints. Surplus, in the context

of this problem formulation, shows where actors are reached (or can be reached) by more than one key player, that is, more than 1 covering. This is an indication of redundancy in key players and actors targeted, which may or may not be desirable. Suppose that a terrorist network is of interest, and the decision maker wants to avoid or minimize the chance that an actor will be contacted by more than one key player in order to avoid suspicion of an 'external' influence. The most straightforward approach to model this situation would be to change all of the covering constraints (≥ 1) to matching constraints ($= 1$). However, depending upon the network topology, and other problem aspects that may be incorporated, this may result in infeasibility. Of course, the analyst can mix and match these conditions for each specific individual, thus meeting the requisite assumptions regarding the sophistication of various target individuals.

On the other hand, redundancy may be required, perhaps to ensure a PSYOP message gets communicated, to include a backup plan should one of the key players change their mind or become unavailable, or to leverage multiple sources to increase the likelihood of a shift in attitudes. To accomplish this, changing the appropriate constraints to (≥ 2), for example, ensures that the corresponding actors are reached by at least 2 distinct key players, or they are themselves a key player and a target of another *kp*-set member. Of course, increasing the right-hand side value further increases the sources of external influence, which may be required for some individuals. Lastly, each of the coverings (by a key player) can be tailored to meet the influence requirements for specific individuals, implying that the right-hand sides need not be identical.

To demonstrate the approach and possible analytical avenues, these mathematical programming techniques are applied to a small example data set.

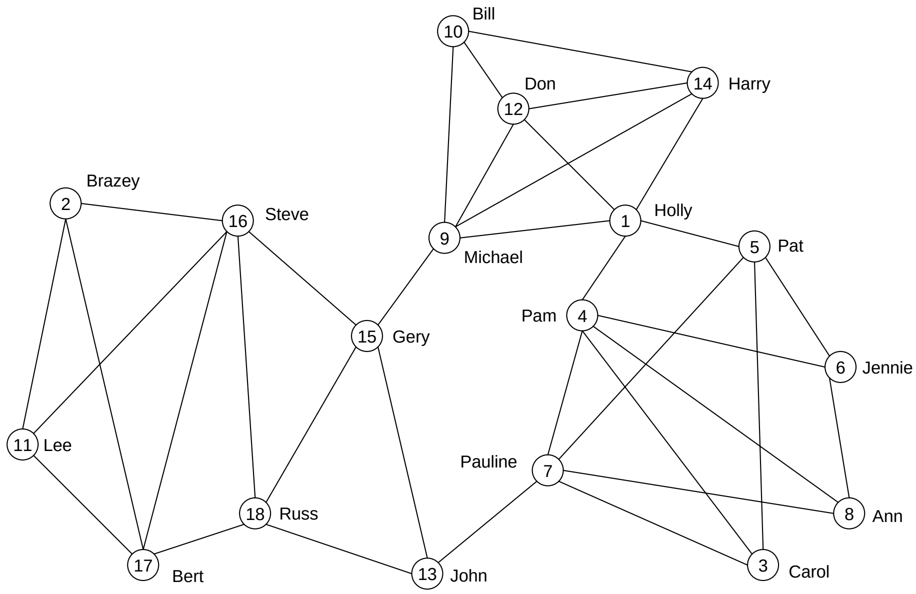


Figure 5.4: Methods Camp Dataset (symmetric) [Borgatti, 2003b]

5.6 Exemplar Case Study

In order to compare results between the heuristic and mathematical programming approaches, one of the data sets provided with the KPP software ('methodscamp') is examined. The network is illustrated in Figure 5.4. As aforementioned, the MP approach is immediately applicable to asymmetric network data. Borgatti's KPP program has not yet been extended to study asymmetric relationships; therefore, the data was made symmetric prior to incorporation into the MP formulations.

5.6.1 Number of Nodes Reached Results

Beginning with the objective of reaching as many actors as possible, given a specified *kp*-set size. Assuming that actors must be within one step from a key player, the minimum number of key players meeting these criteria is 4, the domatic number of this particular graph. There are 18 optimal solutions for (NR1), which

Table 5.2: NR1 Solutions (FNRK1, with $K = 4$)

$\{5, 7, 12, 16\}$	$\{5, 7, 10, 16\}$	$\{5, 7, 9, 16\}$
$\{5, 7, 14, 16\}$	$\{5, 7, 9, 17\}$	$\{6, 7, 12, 16\}$
$\{7, 8, 12, 16\}$	$\{4, 7, 12, 16\}$	$\{7, 8, 9, 16\}$
$\{6, 7, 9, 16\}$	$\{4, 7, 10, 16\}$	$\{\mathbf{4,7,9,16}\}$
$\{4, 7, 14, 16\}$	$\{6, 7, 14, 16\}$	$\{6, 7, 9, 17\}$
$\{7, 8, 14, 16\}$	$\{7, 8, 9, 17\}$	$\{4, 7, 9, 17\}$

is equivalent to (FNRK1) with $K = 4$. The results are shown in Table 5.2. The solutions highlighted in bold in each of tables indicate the single solution generated by the key player software [Borgatti, 2003b]. In some cases, with repeated executions of the heuristic multiple solutions were found for the same problem setting. The mathematical programs developed here (provided in the corresponding appendices) find all optimal solutions if desired.

Multiple optimal solutions not only offer options regarding potential *kp*-sets, but also provide insight into the nature of the solutions. For example, Figure 5.5 depicts a histogram of the number of times a particular player appears within one of the 18 *kp*-set solutions. Actor 7 appears in all 18 optimal solution sets, indicating that if the goal was to reach all actors within one step, actor 7 must be available as a key player. Otherwise, some sacrifices in either the distance assumption or the percentage of population influenced must be considered. Actor 16 appears in 14 of the 18 optimal solutions. When dealing with larger networks, and potentially a much larger number of multiple optima, this approach lends itself to ascertaining the criticality of a given actor and their potential role within a *kp*-set. Combining this technique with other screening criteria facilitates *kp*-set selection.

If the one-step assumption could not be traded off and four actors could not be accessed due to other resource constraints, then the application of the (FNRK1) problem, with reduced values of k is the next logical approach.

From Table 5.3, several interesting observations may be made. As expected from the initial set of results, actors 7 and 16 still play a vital role within a number

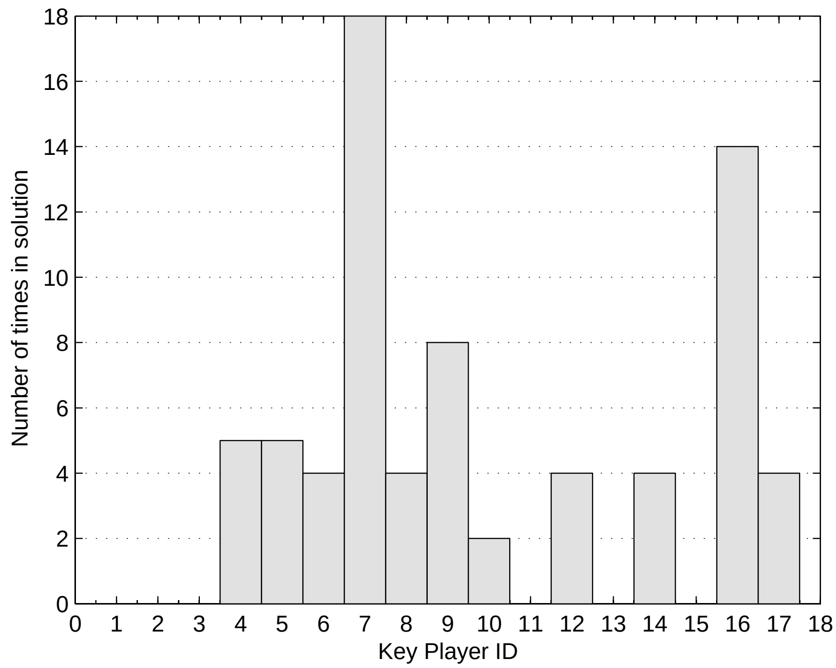


Figure 5.5: Key Player Solution Occurrence

Table 5.3: FNRK1 Solutions (with varying k)

$K = 3$ (17 Actors Reached)
$\{7, 12, 16\}$ $\{7, 14, 16\}$ $\{\mathbf{7,9,16}\}$ $\{7, 9, 17\}$
$K = 2$ (12 Actors Reached)
$\{1, 16\}$ $\{\mathbf{4,16}\}$ $\{7, 16\}$ $\{7, 9\}$
$K = 1$ (6 Actors Reached)
$\{16\}$ $\{9\}$ $\{7\}$ $\{\mathbf{1}\}$ $\{4\}$

of the alternate optima. Additionally, for $K = 3$ the one actor that is missed by all possible solutions is actor 6 (Jennie). This suggests another area of opportunity regarding the tradeoffs that may be offered, based upon data and insights directly gained from examining the MP solutions. If influencing Jennie was not a primary concern, then little solution (and course of action) value is lost due to reducing the *kp*-set size from 4 (reaching everybody) to any one of the *kp*-sets comprised of 3 key players (reaching everybody but Jennie).

Table 5.4: NR2 Solutions (FNRK2, with $K = 2$)

{1,15}	$\{3,15\}$	$\{4,15\}$	$\{5,15\}$	$\{6,15\}$
$\{7,15\}$	$\{8,15\}$	$\{1,16\}$	$\{1,17\}$	$\{1,18\}$

Assuming that the reach between a key player and its assigned actor could extend to two steps, the NR2 program may be applied. The multiple optima are given in Table 5.4. Note that, due to the small size and topology of the given network, (NR2) results are equivalent to (FNRK2) with $K = 2$. Hence, the 2-dominating number for this graph is 2. If only one player could be accessed ($K = 1$), then the only key player under these assumptions is actor 15 (Gery), who can reach 14 other actors within two steps or fewer.

5.6.2 Reciprocal Distance Reach

Turning now to selecting a key player set that maximizes the reciprocal distance reach objective, the (PMED) formulation is initially applied. Note that when $K = 4$, all possible *kp*-sets correspond to the minimum dominating sets of the graph, shown in Table 5.2. This results in a *kp*-set that can reach all other actors within one step ($m = 1$) and therefore provides an objective function that, given these particular assumptions, cannot be improved. Consequently, if an analyst were to explore tradeoffs, it should be in the area of smaller *kp*-set sizes and allowable reach distance. All optimal solutions for (PMED) with $m = 2$ and varying k are provided in Table 5.5. Observe that for $K = 2$, although there are multiple optima from an (NR2) perspective (see Table 5.4), there is only one set $\{1, 16\}$ that optimizes the D_R objective. Of course, the objective function values could be calculated for each of the (NR2) with $K = 2$ solutions to explore second best options. Alternatively, the objective functions for each of these could simply serve as one of several inputs regarding the efficacy or desirability of a given *kp*-set, permitting a multi-objective analysis of the options available.

Table 5.5: PMED Solutions ($m = 2$)

$K = 2$ ($D_R = 0.8333$)
{1,16}
$K = 3$ ($D_R = 0.9722$)
{7,12,16} $\{7, 14, 16\}$ $\{7, 9, 16\}$ $\{7, 9, 17\}$

The (PMEDm) formulation, which permits the analyst to restrict the solution space via reach limitations, essentially only offers a potential means to reduce computational requirements. This is due to decision variables that are (or are not) defined in the math program as a result of $i - j$, m-reach pairs possible. However, if the reach specified, m, does not correspond to any m-dominating set, the mathematical program will be infeasible. In general, the objective function of the (PMED) formulation always selects the shortest path assignment between a key player and its assigned actor(s). Therefore, either solve (PMED) with $m = (N - 1)$, or determine the domatic number in advance for use as input to (PMEDm).

5.7 KPP and Layered Networks

When dealing with the layered network construct, application of the various KPP-2 models may be accomplished in a fashion similar to that discussed for RBAP. That is, each layer could be analyzed independently, with the results across layers compared against one another, or a simple aggregation of all the layers as discussed for RBAP could be performed prior to KPP-2 model implementation. However, the dominating set of a layered graph aggregated by the application of Equations 4.5 or 4.6 is not equivalent to the dominating set of all layers simultaneously.

For example, consider the notional network of three actors in Figure 5.6. If Equation 4.5 is applied to combine the three layers, the aggregated network would take the form shown in Figure 5.7. Clearly, any one of the actors in Figure 5.7 could serve as a minimum dominating set for the graph.

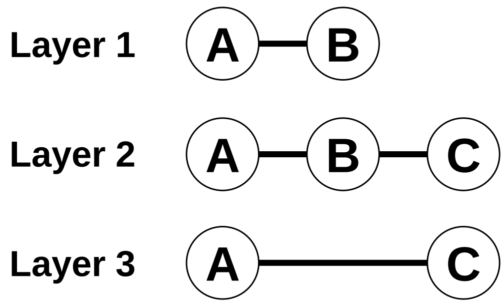


Figure 5.6: Notional 3-Layer Network

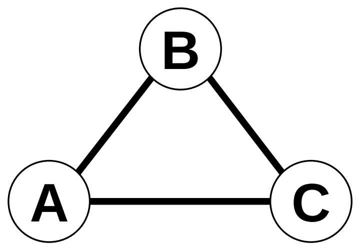


Figure 5.7: Notional Aggregation

However, any single actor cannot dominate all layers simultaneously in Figure 5.6; in fact, any two actors must be included in the minimum dominating set in order for each layer to be dominated. This suggests a new class of problems to address this particular research question. This class is defined as a *multi-layer* or *multi-graph dominating set*, which corresponds to the KPP-2 concept when influence across multiple layers simultaneously is required.

Given a multi-graph comprised of L layers, let G_l, $l = 1, \ldots, L$, represent a given layer l of a graph G such that $G = \{G_1, G_2, \ldots, G_L\}$. Let the corresponding vertices within each layer l be represented by $V(G_l)$, which comprise the superset of vertices $V = \{V(G_1), V(G_2), \ldots, V(G_L)\}$. Edges are defined in a similar fashion such that $E = \{E(G_1), E(G_2), \ldots, E(G_L)\}$. Extending the definitions of the dominating and k-dominating sets offered by Deo [1974] and Penso and Barbosa [2004], respectively, the multi-graph k-dominating set D_m is formally defined as follows.

Definition 10. *Given a dichotomous multi-graph, $G = \{G_1, G_2, \ldots, G_L\}$, with vertices $V = \{V(G_1), V(G_2), \ldots, V(G_L)\}$, the multi-graph k-dominating set is the set of vertices $D_m \subseteq V$ such that $V(G_l) \in D_m$ or is at most k steps away from the nearest vertex in D_m, $\forall\ l$.*

A *minimal multi-graph k-dominating set* is one that satisfies Definition 10 with the minimum number of vertices, the cardinality of which is denoted δ_m. As an initial solution approach, mathematical programming is applied once again. Let $\mathbf{R}_{m_l}$ represent the transpose of the m-reach matrix for layer l. Building upon the original (NRm) formulation as an example, in order to consider all layers simultaneously, the constraint matrix is now comprised of vertical concatenation of $\mathbf{R}_{m_l}$ for each l. This modification is shown in the model (NRmL):

$$
\begin{aligned}
\texttt{(NRmL)} \qquad\qquad & \texttt{Min} \ \sum_{i=1}^{N} x_{i_m} && (5.42)\\
& \texttt{Subject To} \ \mathbf{R}_{m_1}\mathbf{x}_m \geq \mathbf{1} && (5.43)\\
& \mathbf{R}_{m_2}\mathbf{x}_m \geq \mathbf{1} && (5.44)\\
& \vdots \\
& \mathbf{R}_{m_L}\mathbf{x}_m \geq \mathbf{1} && (5.45)\\
& x_{i_m} \in \mathbb{B} \quad \forall\ i.
\end{aligned}
$$

For any given layer, if any particular individual is isolated within that particular context and therefore cannot be reached by any other individual–the potential key players–the result is that column i in the m-reach matrix will be all 0, with the exception of row i, the actor reaching itself. Taking the transpose to form $\mathbf{R}_{m_l}$, which serves as the constraint matrix for problem (NRmL), results in a constraint in the form of $x_{i_{m_l}} \geqslant 1$. This condition consequently requires the inclusion of isolated actors within the dominating, and key player, set.

The other *number reached* formulations could be extended in a similar fashion. For example, if the dominating conditions were not required for all actors over all possible layers, the (FNRK1) formulation may be extended to accommodate not only L layers, but a reach of m steps or fewer between key players and the other actors. Let $y_{i_l} = 1$ if actor i on layer l does not meet the domination criteria, 0 otherwise; and, $\mathbf{y}_l$ is the vector of y_{i_l} for all i. The multi-graph extension to (FNRK1) is (FNRKmL):

$$
\begin{aligned}
\texttt{(FNRKmL)} \qquad \texttt{Min} \quad & \sum_{l=1}^{L}\sum_{i=1}^{N} y_{i_l} && (5.46)\\
\texttt{Subject To} \quad \mathbf{R}_{m_1}\mathbf{x}_m + \mathbf{y}_1 &\geq \mathbf{1} && (5.47)\\
&\vdots \\
\mathbf{R}_{m_L}\mathbf{x}_m + \mathbf{y}_L &\geq \mathbf{1} && (5.48)\\
\sum_{i=1}^{N} x_{i_m} &= K && (5.49)\\
x_{i_m}, y_{i_l} &\in \mathbb{B} \quad \forall\ i.
\end{aligned}
$$

As an additional analysis approach, if one layer was perceived as more important than another with respect to ensuring the dominating criteria are met, a corresponding cost coefficient could be associated with each actor-level combination, y_{i_l}. Therefore, assuming that $K < \delta_m$, such an objective function would minimize the weighted sum of actors missed at each level. Another aspect of this formulation that could represent the criticality or fragility of a given layer is the incorporation of different values of m for each layer. Currently, the formulation in (FNRKmL) assumes a constant m for each layer. This assumption is easily changed based upon analysis assumptions or requirements.

Table 5.6: MP Summary

Problem	Objective	Constraints
Number of Nodes Reached Approach		
NRm	Minimize K	Actors must be within m steps of assigned key player
FNRKm	Minimize actors missed	Actors must be within m steps of assigned key player; K
FNRm	Minimize K	Actors must be within m steps of assigned key player; Can miss at most U actors
DNRm	Distribute workload	Actors must be within m steps of assigned key player; K
Reciprocal Distance Reach Approach		
PMEDm	Minimize $-D_R$	K; Reach of key players is allowed up to a specified $m \in [1, (n-1)]$
FPMED	Minimize $-D_R$	K; Can miss at most U actors
Number of Nodes Reached Approach - Multigraph		
NRmL	Minimize K	Actors must be within m steps of assigned key player for all layers
FNRKmL	Minimize actors missed over all layers	Actors must be within m steps of assigned key player within a given layer; K

5.8 Summary

The mathematical programming approaches offered, summarized in Table 5.6, provide several analytic benefits. These include, but are not necessarily limited to: a guaranteed optimal solution, the incorporation of directed networks, the ability to accommodate valued relations, the ability to discount actors not reachable by external influences, and, in some cases, the ability to encompass multiple dimensions of relationships.

Further modifications permit the fine tuning of MP approaches, incorporating decision maker objectives or other operational requirements and limitations as program constraints. As such, individuals can be designated a key player in advance or

not allowed to be selected at all. Costs to access the various individuals (e.g., military operations, risk, and intelligence resources required) may also be incorporated within any of these formulations. Inclusion of these costs could be accomplished either by a multiple objective linear program or by a modification of the objective functions already discussed.

Multiple optimal solutions may be revealed by post-optimality analysis; identifying these solutions means more viable and effective alternatives for the decision maker. Taking advantage of special constraints, such as the one provided in Equation 5.36, facilitates the enumeration of these alternatives.

Dominating sets and the p-median problem are traditionally difficult problems to solve when the networks are large. Algorithmic improvements, as well as specialized heuristics, for both problems pervade the current literature [cf. Reese, 2005; Grandoni, 2006]. Such efforts offer potentially more computationally efficient alternatives to perform the KPP-2 analyses described in this chapter.

Finally, application of these techniques are not necessarily limited to social networks comprised entirely of individuals and their known relations. Abstraction of this concept to more general networks offers other analytic opportunities. For example, assume the given objective is to influence the citizens within a number of cities of a specified country. Modeling this problem at the citizen level is likely infeasible, due to numerous political and resource limitations. However, modeling the cities themselves as nodes, and perhaps even communities therein, circumvents a number of the computational and data requirements imposed upon such an effort. Consequently, candidates for key players could include media sources, religious or political leaders, or potentially a PSYOP product that can be delivered as influence to a subset of the country's populous. Layered applications are just as flexible. For example, suppose each graph layer represents physical infrastructure. Vertices occurring in multiple layers could represent bridges that facilitate transportation, telecommunications, electric power, and petroleum distribution across a river. Ver-

tices could also represent a mix of people, facilities or processes, mapping two or more social and physical infrastructures together. Considering that KPP-2 seeks to find a minimum set that, in effect, could touch and 'influence' all of these layers, use of this analysis technique could also facilitate vulnerability analysis.

Overall, the conceptual underpinnings of KPP can be found in a number of related operations research problems, as demonstrated by the relationships to existing operations research literature. Developing the linkage of the KPP problem to mathematical programming also lends this problem to the array of heuristic approaches developed for these specific combinatorial problems [cf. Kreher and Stinson, 1999]. These may be particularly useful when analyzing very large social networks in a limited amount of time. However, this particular application, that of influencing a target network from within, is certainly of interest in the current geopolitical climate.

MATLAB code for the (NRm), (FNRKm), (FNRm), and (DNRm) are in Appendices C, D, E and F, respectively. Due to the inherent limitations of the MATLAB solver, as well as the problem difficulty, a more efficient optimization program was required for the (PMEDm) formulation. Appendix G outlines the process used to implement LINGO for this formulation, and presents a small example with data and corresponding solutions.

To this point, the measures and methods described have attempted to analyze the topological structure of networks. The next chapter transitions from a topological focus to one that examines the ties that make up this topology. Hence, the required level of detail in intelligence data is increasing. This data is assumed to capture the nature of the ties between individuals. The questions to be investigated are, 'How do we quantify these ties, and why would we want to?'

VI. Measuring Multiplexity

6.1 Chapter Overview

> The content of relationships is a problem for network analysis. The problem is nicely illustrated in the distinction between naturally occurring relations and analytical relations–the first being the relations in which people are actually involved, the second being the recreation of relations for a network analysis [Burt and Schøtt, 1985]

As Burt and Schøtt point out, there has existed a gap between models of social interaction and reality; to a large extent, this observation remains accurate today. The measures and methods described thus far focus primarily upon analysis of network topology alone. This focus is mirrored by the greater majority of social network measures and sociological studies found within the literature. It is often implicitly assumed that the context of interest within sociological studies is the predominant relation from which the observable network structure has developed. Additionally, links between individuals discovered within these oftentimes specific but still potentially wide-ranging contexts under examination are viewed as homogenous relations, lacking varying degrees of importance or significance when considering interpersonal interactions. Simply reflecting upon our own acquaintances, even within the same context such as work associates or family members, it is likely found that interpersonal relationships vary in many ways.

Consequently, the sociometric representation of those links indicates that either the particular relationship does or does not exist between two individuals, and the dichotomous representation ensues. Therefore, differentiation between "lifelong," "good," and "casual" friends, or other general levels of relationship strengths either cannot be analyzed or are relegated to ad hoc, ordinal measures.

Quantifying how these relationships vary, capturing the very nature of an organization's interpersonal ties, is the objective of this chapter. The questions to be investigated are: How do we effectively quantify these ties, and why would we want

to? Several means to answer the first part of this question are posed. The rationale for attempting such a measurement is to both improve upon previous network flow methodologies and to provide a more accurate representation of the nature of the relations. Previous works have merely posited the existence and attainment of such values [e.g., Stephenson and Zelen, 1989; Freeman et al., 1991; Yang and Knoke, 2001], while others [e.g., Renfro, 2001; Clark, 2005] have presented new means to estimate what may be viewed as tie strength.

The next sections first clarify the concept of tie strength, as several meanings and definitions have been previously offered within the literature. Previous works that have suggested means to measure this phenomenon are reviewed; and, new methods are offered that are both based upon predominant sociological theory related to the concept of tie strength and amenable to implementation despite limited available information. One such method implements a decision theoretic model of tie strength. This last approach permits extensions of several weighting concepts (discussed earlier in Section 2.5.5.1) that offer a means to capture the effects of structural change upon a network's ties.

6.2 Tie Strength

In the sociological literature at least two definitions of strength of a personal tie may be found. Granovetter defined the strength of a tie as "a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie" [Granovetter, 1973, pg. 1361].

Interestingly, the converse of strong ties, weak ties, are composed of casual or intermittent relationships and are potentially strong themselves. The strength of a weak tie lies in its ability to bridge communication or influence between two or more distinct groups, or promote diffusion of influence and ideas between them

[Granovetter, 1973, pg. 1363-7]. In this latter case, the strength lies within potential opportunity for the exchange of information, services, or influence that is otherwise non-existent within one's own immediate sphere of influence or stronger acquaintances.

Although weak ties may be of interest with regards to identifying individuals that offer critical and highly-skilled services, an ability to craft explosives or biological weapons, for example, for the purposes of this research, reference to a strong tie implies the former definition. Hence, it is assumed that relationships sharing multiple contexts result in the significant bonds of trust required of a non-cooperative, and particularly a terrorist, network. Interaction within multiple, and not necessarily just social contexts, is a phenomenon referred to a multiplicity, which recognizes that virtually all "naturally occurring [relations are] a bundle of different interaction elements" [Burt and Schøtt, 1985, pg, 288]. As shown by Levin et al. [2002], "strong ties promote effective knowledge because they tend to be trusting ones" [Levin et al., 2002, pg. D2]. Trust–"that quality of the trusted party that makes the trustor willing to be vulnerable"–is assumed to play a key role in binding the network membership together [Mayer et al., 1995, pg. 712].

A recent survey by Hite investigated a number of sociological studies that characterized interpersonal ties. Note the conceptual variety illustrated within Table 6.1 and that none of these deal with non-cooperative networks. The specific contexts that significantly contribute to the strength of an interpersonal tie between two members of a non-cooperative network remains an open research question. This is primarily due to the lack of previous, unclassified, investigations of terrorist organizations (other than Renfro [2001] and Clark [2005]), as well as the multitude of underlying motivations and cultural phenomena that result in the formation of these organizations. For example, the Islamic basis for action and overall intentions and organizational objectives of Al Qaeda are presumably much different than those of the Columbian narcoterrorism organizations National Liberation Army (ELN) and

Table 6.1: Tie Characteristics [Hite, 2003, pg. 14]

Concepts (Source)
Affect; philos; passions (Granovetter, 1985; Krackhardt, 1992; Uzzi, 1999)
Frequency or frequent contact (De Burca et al., 2001; Granovetter, 1985)
Reciprocity (Granovetter, 1985; Portes and Sensenbrenner, 1993; Powell, 1990; Uzzi, 1999)
Trust; enforceable trust (Portes and Sensenbrenner, 1993; Powell, 1990; Uzzi, 1996)
Complementarity; accommodation and adaptation; indebtedness or imbalance; collaboration; transaction investments; strong history; fungible skills (Powell, 1990)
Expectations; social capital; bounded solidarity (Portes and Sensenbrenner, 1993)
Lower opportunistic behavior (Provan, 1993)
Density (Staber, 1994)
Maximize relationships over organization (Powell and Smith-Doerr, 1994)
Fine-grained information transfer; problem solving (Uzzi, 1996)
Duration; multiplexity (De Burca et al., 2001; Uzzi, 1999)
Diffusion; facilitation (MacLean, 2001)
Personal involvement; low formality (few contacts); connectedness (De Burca et al., 2001)

the Revolutionary Armed Forces of Columbia (FARC) [Berry et al., 2002]. Different world-views and goals correspondingly result in different opinions, from the perspective of the organization's members, of what is or is not important with regards to interpersonal ties.

Previous efforts that have attempted to identify the significant components of tie strength are discussed in Marsden and Campbell [1984] and Carroll [2006]. Both used canonical correlation analysis to identify general relationships between multiple tie characteristics and multiple aspects of the resulting strength. Carroll investigated the types of alliances formed among financial and commercial institutions as a result of multiplex relationships. Although Carroll's concept and findings may be of interest

with regards to inferring strength between geographically separated terrorist cells that have little or no direct contact (other than through the Internet), the exact methodology used by the authors is not immediately transferrable due to the survey nature of the data. The specific findings of Marsden and Campbell [1984], however, offer some insight that may be leveraged in a decision theoretic model of tie strength. The next sections discuss the interrelated concepts of tie strength, social distance, and social closeness, as well as previous efforts attempting to measure them.

6.2.1 Distance, Closeness, & Strength

Perhaps the first sociological attempt to measure or quantify interpersonal relationships was due to Bogardus [1925], who defined social distance as “the degrees and grades of understanding and feeling that persons experience regarding each other” [Bogardus, 1925, pg. 299]. Bogardus’s intentions were to “chart the character of social relations” between various ethnicities [Bogardus, 1925, pg. 299]; his work continues to be studied and used to monitor the longitudinal trends in race relations [cf. Parrillo and Donoghue, 2005].

Conceptually, social distance is effectively inversely proportional to tie strength. Bogardus attempted to measure social distance by asking individuals about their propensity to include other races in their social circles, at varying degrees of intimacy. As seen in Table 6.2, social distance as measured by Bogardus is predicated upon a Guttman scale, where, for example, responding affirmatively to condition 1 implies the same response for conditions 2 through 6. This corresponds to the smallest social distance, and therefore the strongest potential contact strength. Alternatively, responding affirmatively to condition 2 implies the same response for conditions 3 through 6, and so forth. An affirmative response to condition 7 implies that no tie is sought, in this setting, to another race, therefore resulting in significant social distance.

Table 6.2: Social Distance Measurement [Bogardus, 1925, pg. 301-3]

Willingly admit...	Social Contact Distance	Contact Strength
1. To close kinship by marriage	0	Strong
2. To my club as personal chums	1	
3. To my street as neighbors	2	
4. To employment in my occupation in my country	3	↓
5. Citizenship in my country	4	
6. As visitors only to my country	5	Weak
7. Would exclude from my country	6	Non-existent

The Guttman scale approach attempts to incorporate the perception that "different interpersonal processes occur at different stages of a relationship's development" [Friedkin, 1990, pg. 240]. However, Bogardus's work is not only clearly uniplex, but is reliant upon survey data collected by open and seemingly honest participants. Friedkin's application of this concept does involve multiplexity to an extent, incorporating claims of "friendship," "help seeking," and "frequent discussion" among participants [Friedkin, 1990, pg. 240-1]. Unfortunately, this too is reliant upon truthful survey responses and is primarily qualitative in nature. In fact, due to the difficulty of mapping a measure to the strength of an interpersonal tie, the literature is dominated by qualitative analysis in the subject [cf. Jack, 2005; Granovetter, 1973; Pabjan, 2005, among others].

Recalling the relationship between social distance and tie strength, Renfro's concept of social closeness is based upon (psychological) distance. Closer individuals had less distance between them, and therefore shared stronger interpersonal ties. Given a decision theoretic measurement of an individual's current psychological state, Renfro's measure of strength was derived by taking the difference of two individuals' scores [Renfro, 2001, pg. 178-81]. Although the decision theoretic model proposed by Renfro is well founded in psychological and sociological theory, the data collection

efforts and cultural expertise required may prove overwhelming when dealing with large networks of individuals engaged in surreptitious activities.

More recently, Clark utilized the information centrality measure developed by Stephenson and Zelen [1989] to formulate a matrix of pair-wise (interpersonal) influence measurements, defined by

$$\mathbf{W} = w_1\mathbf{I}_1 + w_2\mathbf{I}_2 + \ldots + w_n\mathbf{I}_n : \sum_{i}^{n} w_i = 1. \tag{6.1}$$

This measurement is based upon a weighted sum of the matrices that would normally be used to calculate information centrality for the network's actors within the respective layers. Instead, the linear combination of matrices, one matrix for each contextual layer, are combined and multiplied again by a coefficient that serves as a proxy for individual-specific influence, denoted $\mathbf{e}$. Inclusion of the influence attributed to specific individuals, based upon actor attributes, essentially induces asymmetry in the matrix $h_{ij} \in \mathbf{H} = \mathbf{W_{ij}e_i}$. The elements $h_{ij} \in \mathbf{H}$ are then used for a variety of analyses, including network flow formulations [Clark, 2005, pg. 3-34].

Clark's measure essentially attempts to capture the potential amount of influence that one person may impose upon another. This is analogous to Renfro's application of social closeness to arc capacity within a network flow formulation of the social network. There are, however, potential issues associated with this approach, depending upon the inherent structure of the data. For example, for a given layer l, the (i, j)th entry within the $\mathbf{I}_l$ matrix captures the "information in the combined path" between individuals i and j [Stephenson and Zelen, 1989, pg. 12]. Consequently, despite the lack of an existing, direct relationship between any two individuals in l, the (i, j)th entry, $i \neq j$, may be non-zero. The weighted combination approach ultimately separates the contextual layers; thus, despite indirect connections between actors through inter-layer connections, the relationships between those individuals are mathematically ignored.

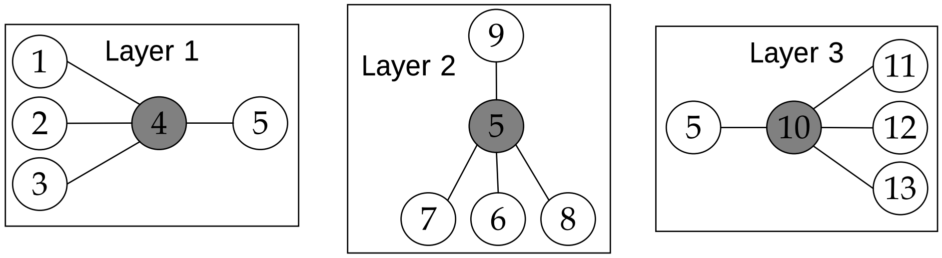


Figure 6.1: Notional Network Layers

Consider the network shown in Figure 6.1, which is comprised of 13 individuals within 3 different layers. The fact that information centrality attempts to measure the information flowing along all possible paths within a network suggests that globally there should be some information exchange, albeit indirectly, between actors 1 and 13, for example. Since actors 1 and 13 do not communicate within the same layer, applying information centrality to each layer as a stand-alone network and then aggregating the results via Equation 6.1 is mathematically contrary to the premise of the measure proposed by Stephenson and Zelen [1989].

This issue does not discredit the approach entirely, even though the information centrality measure due to Stephenson and Zelen [1989] is potentially flawed in itself (See Appendix H). A remedy could include using a weighted combination of layers as input to the information centrality calculations. For example, given the layers in Figure 6.1 and assuming the weights were 0.5, 0.3, and 0.2 for layers 1, 2, and 3, respectively, the network shown in Figure 6.2 would serve as the input to this measure.

Assuming that this approach yields a connected graph and that the resulting weight values are substituted for the traditionally dichotomous representation of the graph, the necessary conditions for the information centrality measure calculations are met. Larger weights subsequently induce a corresponding bias toward relationships, and the paths they form, within the information centrality results. Applying

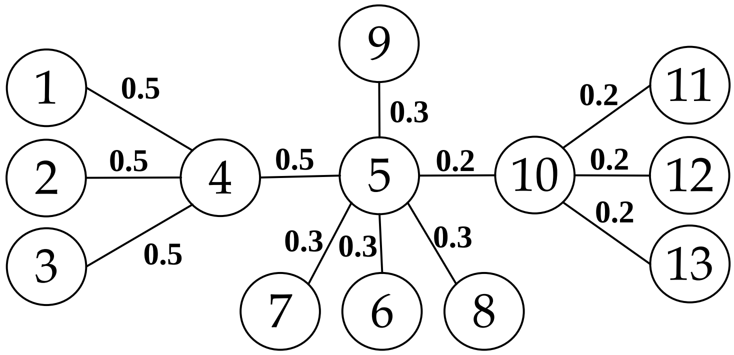


Figure 6.2: Weighted Combination of Layers

the information centrality measure to the entire, connected graph also ensures that potential information between indirectly connected actors is measured.

Distance, closeness, and strength are similar, if not identical constructs of interpersonal relationships. The closer two individuals are, the smaller the social distance and the stronger the interpersonal tie. For this research, the nomenclature of *tie strength* was chosen simply for its conceptual clarity–the direct relationship between tie strength and the value for its measure. For the purposes of this research, tie strength is formally defined as follows.

Definition 11. *The strength of an interpersonal tie between individuals i and j, denoted $s_{ij} \in [0, \mathbb{R}^+]$, measures the degree of trust and shared understandings between two people, relative to all other contacts within the appropriate social contexts shared among the network of interest. The stronger (weaker) the interpersonal tie, the greater (smaller) the value of s_{ij}. Actors i and j who have no direct, interpersonal relationship have a strength of zero.*

This is very similar to Renfro's definition of *social closeness*, and formalizes the topological aspect of Clark's holistic interpersonal influence measure, as well as the notional concept of value placed upon a tie in Freeman et al. [1991]. Consequently,

the models proposed in the next section could also serve as arc capacities in a network flow model of a social network.

However, another option could entail viewing the strength of a tie as the cost associated with interpersonal communications. As Jack observed, information and support gained from strong ties has several benefits, it is "more trustworthy because it is richer, more detailed and accurate; it is usually from a continuing relationship and so in economic terms it is more reliable" [Jack, 2005, pg. 1236]. Strong ties, then, may permit more efficient or less costly communication paths, particularly when the information may be counter to the organization's objectives or potentially detrimental to those who promulgate the information to others. If relating tie strength to cost vis-a-vis Jack [2005] was of interest, the cost for individual i to communicate with individual j, c_{ij} would be inversely proportional to s_{ij}. Stronger ties would have smaller costs. Non-existent ties would effectively have an infinite cost, as shown by Equation 6.2, and would therefore not contribute to the objective function of a network flow formulation, or any other linear program, with a feasible region.

$$\lim_{s_{ij} \to 0^+} \frac{1}{s_{ij}} = \infty \tag{6.2}$$

Several methods to ascertain interpersonal tie strength are proposed, all of which assume that limited information regarding the nature of the ties exists, and expert opinion is able to determine which social contexts significantly contribute to the cohesion and continuation of the organization under study.

6.3 Models of Tie Strength

The objective of this chapter is to quantitatively characterize the strength of interpersonal ties. This characterization must be possible with limited information regarding the social interactions of the individuals, primarily due to their own requirements of secrecy, deception, and detection avoidance. Making the mathemat-

ical connections between these individuals, however, is predicated upon the contexts within which they operate, train, and build cohesive and trusting relationships [Marsden and Campbell, 1984, pg. 488]. Such contexts could comprise the layers depicted in Figure 1.1.

As Renfro noted, the simplest method of "counting the number of arcs incident to the individuals involved, . . . or the number of times pairs of individuals communicate in a fixed time period" [Renfro, 2001, pg. 22]. Frequent interactions, and the time required to achieve them, have been suggested to contribute to tie strength [Granovetter, 1973, pg. 1362]. However, Marsden and Campbell found that

> the use of frequency as a measure of strength will tend systematically to overestimate the strength of ties between persons who are neighbors or co-workers, while the use of duration as a measure of strength will overestimate the strength of ties between relatives [Marsden and Campbell, 1984, pg. 499].

Within the next sections, methods for dealing with multiple layers and using them to estimate interpersonal tie strength are proposed. Although the first method is computationally attractive, a few inherent conceptual disadvantages are discussed. A means to deal with such challenges is addressed by a decision theoretic model of tie strength. This model is mathematically similar to the simple aggregation methods discussed in both Chapter II and the following section; however, it incorporates the relative importance–from the individual's perspective–of the various layers or contexts that comprise the 'bundle of associations' forming the relationship.

6.3.1 Simple Aggregation

There is a general agreement within the literature that actor similarity directly correlates to tie strength [Granovetter, 1973, pg. 1362]. Although this phenomenon is traditionally associated with similarity among individuals' attributes, referred to as homophily, one could extend this concept to similarity of contexts, assuming that similar interests and attributes contributed to the shared contexts. Therefore, ex-

tending the shared similarities among attributes, a the Jaccard similarity coefficient may be applied to a network's multiple contexts. This results in a single network with arc values representing the interpersonal tie strength relative to all other individuals within the network.

Given two sample sets A and B, the Jaccard similarity coefficient Yin and Yasuda [2005, pg. 474], $J(A,B) \in [0,1]$, is given by

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \tag{6.3}$$

This measure is 0 (1) if the two sets are completely dissimilar (similar). Assuming that relationship strength is enforced by sharing not only by common contexts that comprise the interpersonal relationship between A and B, but by shared contacts within those contexts, a proxy for strength can be derived by applying the Jaccard measure.

Let E_l denote the set of all edges in layer $l \in L$. Further, let $E_l(i) = 1$ if actor i is adjacent to a given edge $\in E_l$, 0 otherwise. The edge set for a given actor A is then defined as $\{E_1(A)|E_2(A)|\ldots|E_L(A)\}$. The edge set for actor B is defined in the same manner. Note that for a given actor i, these sets correspond to the ith row of a node-edge adjacency matrix. Let $\mathbf{S}_J$ denote the Jaccardian similarity matrix where $S_J(A,B) \in S_J = J(A,B), A \neq B$. The values within this matrix then serve as a proxy for the strength of interpersonal ties. (See Appendix I for the MATLAB code that calculates this measure.)

Consider the complete graph of 5 individuals in Figure 6.3 and its corresponding node-edge adjacency matrix, $\mathbf{E}$, in Equation 6.4, representing a notional layer or terrorist cell. Since all individuals have the same number of connections, shared among all others, the tie strengths are all equivalent. Interestingly, for complete graphs, larger networks result in smaller values for the similarity coefficients. In general, using Equation 6.3, the tie strength for a complete graph, and single layer,

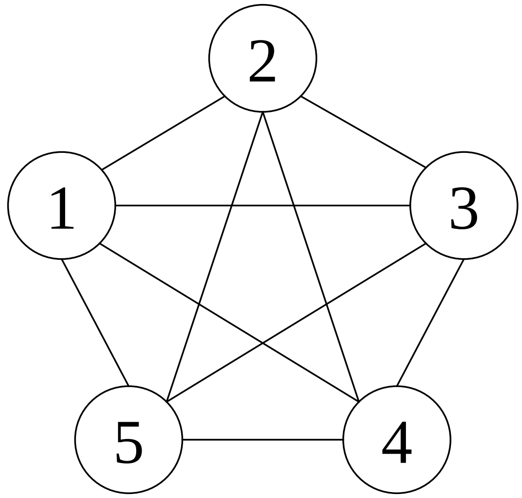


Figure 6.3: Five-Actor Complete Graph

is shown by Equation 6.5. The resulting tie strength matrix corresponding to Figure 6.3 is shown in Equation 6.6.

$$\mathbf{E} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \tag{6.4}$$

$$s_{ij=} \left|i \cap j\right| / \left|i \cup j\right| = 1/\left[(N-1)+(N-1)-1\right] = (2N-3)^{-1} \tag{6.5}$$

$$\mathbf{S}_J = \begin{pmatrix} 0 & 0.1429 & 0.1429 & 0.1429 & 0.1429 \\ 0.1429 & 0 & 0.1429 & 0.1429 & 0.1429 \\ 0.1429 & 0.1429 & 0 & 0.1429 & 0.1429 \\ 0.1429 & 0.1429 & 0.1429 & 0 & 0.1429 \\ 0.1429 & 0.1429 & 0.1429 & 0.1429 & 0 \end{pmatrix} \tag{6.6}$$

Larger complete networks would yield similar results but with smaller values for tie strength. Thus, as an individual's time, cognitive demand, and attention becomes more dispersed, the more detrimental the effect upon tie strength. This can change, however, when a bias between one or more individuals is introduced. Such a bias, while using this approach, may be accomplished by incorporating multiple layers. For example, suppose the initial graph in Figure 6.3 represented operational ties of some sort. Consider another context or layer capturing the familial relationships among the five members. Suppose that members 1, 2 and 3 are all related in a meaningful positive familial way. All other individuals have no common, familial ties. The **E** matrix is shown in Equation 6.7. Note that the first 10 columns correspond to the first, operational layer; the last 3 columns correspond to the familial layer.

$$\mathbf{E} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \tag{6.7}$$

$$\mathbf{S}_J = \begin{pmatrix} 0 & 0.2000 & 0.2000 & 0.1111 & 0.1111 \\ 0.2000 & 0 & 0.2000 & 0.1111 & 0.1111 \\ 0.2000 & 0.2000 & 0 & 0.1111 & 0.1111 \\ 0.1111 & 0.1111 & 0.1111 & 0 & 0.1429 \\ 0.1111 & 0.1111 & 0.1111 & 0.1429 & 0 \end{pmatrix} \tag{6.8}$$

The results in Equation 6.8 depict the shift of tie strengths due to increased similarity (or elements in common) among actors 1, 2, and 3. Accordingly, the tie strengths between these actors and the others not as well connected decreased; such phenomena is mathematically due to the changes in similarity and practically due to the finite cognitive and time resources available to maintain varying relationships.

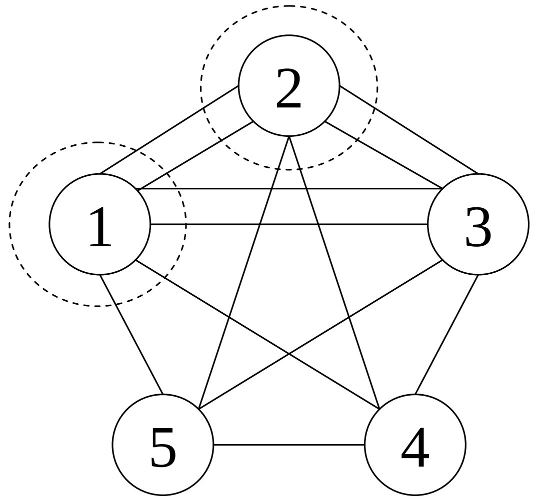


Figure 6.4: Strength Example s_{12} with Five-Actor Multigraph

This effect is also illustrated graphically in Figure 6.4, which shows the two layers superimposed upon each other to form a hypergraph. As an example, consider the tie strength between actors 1 and 2, highlighted by the dashed circle. The Jaccardian measure is essentially the total number of links shared between actors 1 and 2, divided by the total number of links emanating from both actors, or $s_{12} = 2/10 = 0.2000$.

Along this line of thought, the measure values also correspond to ties generally described as weak, connecting separate subgroups [Granovetter, 1983], or as bridges [Brass, 1995]. For example, consider the notional network having two apparent subgroups with two links between them in Figure 6.5. The ties between actors 2 and 6, and 4 and 10 are among the lowest values for tie strength. In general, the more opportunity two individuals have to focus on each other, instead of their other social contacts, the stronger the resultant tie.

Of course, this approach implicitly assumes that each layer or context contributes equally to the strength of a tie. In reality, some contexts may be more meaningful than others, as perceived by the individuals' culture, organizational history and objectives, and world view. This suggests that if experts can determine to

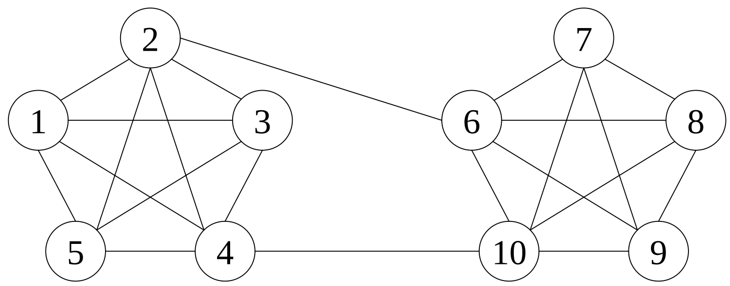


Figure 6.5: Notional Network with Connected Subgroups

what extent the significant components of an organizations ties are, as well as their relative contribution to tie strength, a weighted approach similar to that suggested by Clark may be more appropriate. This approach is discussed next.

6.3.2 Decision Theoretic Approach

The decision theoretic model proposed here requires the following underlying assumptions: (1) information characterizing non-cooperative networks is available; (2) this information is comprised of, or can be broken down into, multiple layers or contexts that are perceived to significantly contribute to tie strength; and, (3) the degree to which each layer contributes to the trust and understanding between two people as defined in Definition 11 is known or can be effectively estimated.

The model is based upon the construct of tie strength due to Granovetter [1973, pg. 1361], which suggests that time, intensity of emotions invoked by the relationship, level of intimacy among the two individuals, and the exchange of or reliance upon another's services may potentially capture the strength of a tie [Granovetter, 1973, 1361]. The corresponding model is shown in Figure 6.6.

"Reciprocal Services" are currently not incorporated within the value model for tie strength. This component could be included, assuming a formal definition and corresponding measure is developed. However, this particular aspect of tie strength deals with asymmetric or negative ties [Granovetter, 1973, pg. 1361]. Asymmetry

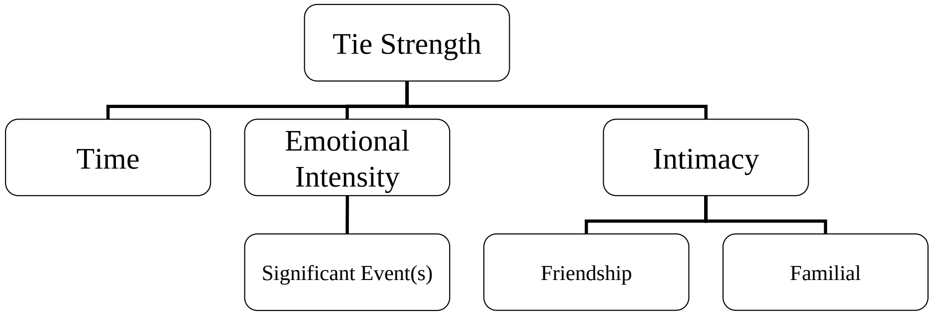


Figure 6.6: Value Model of Tie Strength

of tie strength, especially when considering the relationship of strength to either arc capacity or cost within a network flow formulation, is hypothesized to be a function of individual characteristics, rather than the underlying composition of an interpersonal tie. As such, asymmetry arising from differences between individuals is discussed in Chapter VII. Negative ties are assumed beyond the scope of this research but are certainly of future interest within the context of applying information operations against a target network.

The evaluation measures for this model are summarized in Table 6.3. Of the two potential indicators of tie strength, frequency of interaction and time spent within a relationship, Marsden and Campbell found that the correlation between the (self-assessed) tie strength and time spent was the most appropriate, as the frequency of interaction tends to be confounded or conditional upon the nature of the relationship. Consequently, the first component of this model captures this temporal aspect. The measure is specifically defined as "the time elapsed since the initial observation of active participation in any significant event." This data for a given pair of individuals is essentially captured when that relationship is discovered– using either that specific point in time as the basis of the measure or information about the tie that explicitly states when the tie was formed. The amount of time is then compared to the maximum amount of time elapsed for the oldest known tie

Table 6.3: Evaluation Measures for Tie Strength

Title	Measure Unit	Measure Type	Lower bound	Upper bound
Time	Time elapsed since the initial observation of active participation in any significant event	Time (Linear)	0	$\max(t_{elapsed})$
Emotional Intensity				
Significant Events	Have the individuals actively participated in one or more events or are associated with each other within a given context, either of which is believed by domain experts to significantly contribute to the trust and relational bond between the two participants	Binary	0	1
Intimacy				
Friendship	Have the individuals self confirmed a level of friendship?	Binary	0	1
Familial	Are the individuals family members?	Binary	0	1

within the network, over all layers. This allows the comparison between the most senior and new members of the group.

This approach measures the time spent within a tie and assumes that increased time within a relationship correlates to increased tie strength [Marsden and Campbell, 1984]. Although the measure currently assumes a simple linear relationship between time spent and tie strength, other forms of value functions could be used. For example, an s-curve as shown in Figure 6.7 could be used to represent probationary or indoctrination periods enforced by the target network, or some other period of time generally required by the organization of interest regarding who they may trust and how much.

The domain of this function is bounded above by the length of the oldest known relationship. In addition, given a specific pair of individuals, if multiple contexts are shared among their relationship within the Emotional Intensity or the Intimacy

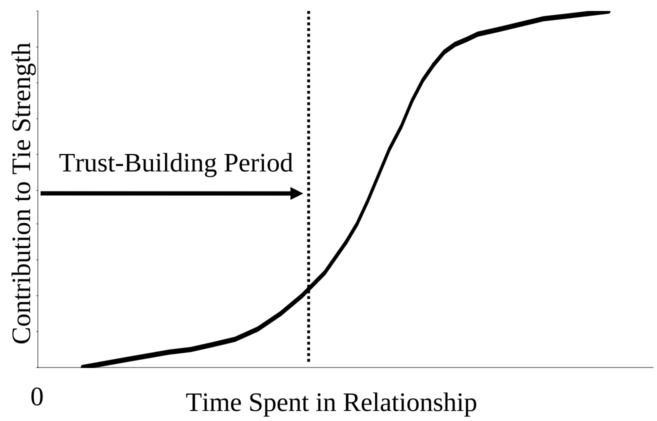


Figure 6.7: Time and Tie Strength

objectives, it is currently assumed that the contribution of time to tie strength is based upon the oldest context–the one within which the relationship began.

Recalling that the objectives and their corresponding measures must be independent for an additive value model, the connection inherent between time spent within a relationship and its existence has the potential to be problematic. However, given two pairs of individuals $A-B$ and $C-D$ with relationships composed of identical contexts, variation of tie strength among them is assumed to be explained by the amount of time those pairs of individuals have had to develop the relationship and concomitant levels of trust [Marsden and Campbell, 1984; Granovetter, 1973, 1983; Levin et al., 2002]. In this sense, the objectives may then be viewed as mathematically independent.

The Emotional Intensity and Intimacy objectives attempt to ascertain tie strength due to one or more significant, bonding events and friendship or familial ties, respectively. These types of contexts that may comprise an interpersonal relationship essentially form the different layers within the layered-network paradigm. Significant bonding events are those that not only bring individuals together, but induce trust via the knowingly joint involvement in a stressful situation, either physically or in the form of cognitive dissonance [Downs, 2006, pg. 3-6]. Examples of such

events could include indoctrination, training, attending and actively participating in educational forums exploring activist or extremist ideologies contrary or significantly different from generally accepted practices, and so forth.

Determining which contexts play the most significant role in tie strength is not a straightforward task, as they are likely dependent upon the network of interest. The network's goals and objectives, methods of recruitment and training, and any other potential layer of contributing relations must also be viewed from their perspective rather than our own. The secretive nature of the organizations of interest may indirectly determine which contexts are important, based upon the contexts and connections inherent within the observational data.

Overall, the value model approach is similar to a simple weighted summation of multiplex relations, further moderated by the age of the relation. This is amenable to situations where very limited data exists, in both quantity and quality. In addition, assuming that the presence of an additional context increases tie strength, then according to Sarle [1995], the binary variables connoting the presence or absence of layers between individuals are at least ratio [Sarle, 1995]. Considering that the contribution of time to the strength of a tie is also ratio, the weighted combination required to ascertain a single value is ratio as well. Consequently, this measurement approach is mathematically appropriate for use as either an arc capacity or a cost per unit flow within a network flow model of the social network. The next logical task is how to weight each of these measures.

6.4 Weighting & Tie Strength

As discussed in Section 2.5.5.1, a variety of weighting techniques exist. Although indifference measurement appears to be the most theoretically sound approach to weighting value models over that of numerical estimation methods, both methods continue to be applied and debated within the current literature.

Note that the overall purpose of the decision theoretic model presented in this chapter is to estimate the strength of interpersonal ties from the individuals' perspectives, not our own. Consequently, eliciting weights from members of a non-cooperative network is assumed to be problematic and unreliable. What can be done, however, is to examine the underlying structures and contexts that have led to the network data discovered to date. Based upon such a review, further hypotheses may be offered regarding which context or contexts actually resulted in the observed topology, and therefore play significant roles in how the members initiate and sustain their clandestine relationships.

For the purposes of this research, it is assumed that subject matter experts, cognizant of the cultural, sociological, and operational aspects of the target network, are able to provide at least initial estimates of the model weights. This may be accomplished via the numerical estimation techniques shown in Table 2.14. As an alternative, if one also assumes that tie strength derived from multiple contexts is also a function of the number of ties existing within that context, dynamic weighting approaches may be applied. The following definitions explain the overall model, shown in Equation 6.9.

$$
\begin{aligned}
s_{ij} &= \text{Estimated strength of interpersonal tie between } i \text{ and } j \\
T_{ij} &= \text{Time elapsed of the first context shared by } i \text{ and } j \\
w_T &= \text{Weight for the } \textit{Time} \text{ component, } T_{ij}\text{, of the model} \\
SEl_{ij} &= 1 \text{ if } \exists \text{ a relationship between } i \text{ and } j \text{ within context } l\text{; 0 otherwise} \\
w_{SEl} &= \text{Weight for } SEl_{ij} \\
w_{EI} &= \text{Weight for the } \textit{Emotional Intensity} \text{ component of the model} \\
FR_{ij} &= 1 \text{ if } \exists \text{ an acknowledged friendship between } i \text{ and } j\text{; 0 otherwise} \\
w_{FR} &= \text{Weight for } FR_{ij} \\
FA_{ij} &= 1 \text{ if } \exists \text{ a familial relationship between } i \text{ and } j\text{; 0 otherwise} \\
w_{FA} &= \text{Weight for } FA_{ij} \\
w_I &= \text{Weight for the } \textit{Intimacy} \text{ component of the model}
\end{aligned}
$$

$$s_{ij} = w_T T_{ij} + w_{EI}\left(\sum_{l=1}^{L} w_{SEl} SEl_{ij}\right) + w_I\left(w_{FR} FR_{ij} + w_{FA} FA_{ij}\right) \tag{6.9}$$

If dynamic weighting was of interest for the weights associated with network layers (w_{SEl}, w_{FR}, w_{FA}), at least two approaches are available. The first, and most straightforward approach, is to assume that the propensity of relations in a given layer l is directly proportional to that layer's contribution to tie strength. For example, suppose E_l is the number of edges (relations) in a given layer $l \in L$ and E_L is the sum of all edges in the network. The relative weight for a given layer l is given by $w_l = E_l/E_L$, or the percent of all ties attributed specifically to the lth layer. As the network evolves and changes over time, so would the weights; since these are percentages, this approach also maintains a normalized weight set.

A second approach first requires initial estimates of relative weights for each layer w_l. Next, define $\rho_l(t)$ as the ratio of the number of edges in a given layer l to the total number of edges in the network data, at a specified time t:

$$\rho_l(t) = \frac{E_l\,(t)}{E_L\,(t)}. \tag{6.10}$$

Let δ_l denote the relative change in in the network due to the edges within layer l, over a time period from t to $t+1$. The value, $\delta_l \geq 0$, is

$$\delta_l = \frac{\rho_l(t+1)}{\rho_l(t)} = \left[\frac{E_l\,(t+1)}{E_L\,(t+1)}\right]\left[\frac{E_L\,(t)}{E_l\,(t)}\right]. \tag{6.11}$$

The weights at time $t+1$, adjusted for relative changes in the composition of the network layers, denoted ω_l, is

$$\omega_l = \frac{\delta_l w_l}{\sum\limits_{j=1}^{L} \delta_j w_j}. \tag{6.12}$$

Note that combination of Equations 6.11 and 6.12 have several desirable properties. First, if no changes are evident within the layers, the original weights specified by the subject matter experts are still applicable and remain constant. Second, if a layer, and its associated edges is eliminated for some reason, the weight for that layer goes to zero; this matches the intuitive result with regards to the now non-existent layer's contribution to tie strength. Finally, relative allocations among weights that remain unchanged over time also remain constant; this preserves the tradeoffs originally estimated by the subject matter experts and is . For example, suppose the original weight matrix is given by $\mathbf{w} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \end{bmatrix}$. If the edges associated with layer two are all removed, the adjusted weight vector is $\omega = \begin{bmatrix} 0.71429 & 0 & 0.28571 \end{bmatrix}$. Note that the ratios between the first and third weights ($0.5/0.2 = 0.71429/0.28571 = 2.5$) remain constant. This holds true for any pair of weights that correspond to unchanging layers over time.

6.5 Summary

This chapter reviewed some of the details, advantages, and disadvantages of previous efforts seeking to measure interpersonal relations. A formal definition of interpersonal tie strength is offered to provide some conceptual clarity, and two models were proposed to facilitate its measurement. Both models assume that limited information is available due to the clandestine and adversarial nature of the networks of interest to the U.S. Government.

The first model applied the Jaccardian similarity coefficient to estimate the strength of a tie between two individuals. This coefficient takes into account both the context(s) between two individuals, as well as the cognitive demand upon them due to the other ties they currently maintain. The second model takes a decision theoretic approach to tie strength, leveraging the findings within sociological literature describing the components of interpersonal tie strength. Due to the inherent nature of the target network and its associated data, direct and dynamic weight-

ing techniques using classical numerical estimation techniques are suggested. The mathematical properties, either in a static or dynamic sense, are attractive for use in mathematical programs due to their ratio scale. As seen in previous research efforts, the measures of tie strength could be used as arc capacities within a network flow formulation. However, the potential relationship between tie strength and the 'cost' of social interaction also suggests that the inverse of tie strength could serve as the cost per unit flow along an arc within a similar network model. The next chapter continues the development of translating observational data for use within mathematical programming applications, focusing upon the concepts and applications of gains, losses, and thresholds.

VII. Gains, Losses, and Thresholds

7.1 Chapter Overview

In light of the complex and elusive terrorist networks that have become of more prominent interest in U. S. since the September 11th attack, it is not only important to know the enemy but also to know the individuals with whom they interact: their friends, enemies, confidants, relatives, classmates and collaborators. The information garnered by uncovering and analyzing such networks offers a potential means to generate, and possibly evaluate, courses of action that shape the intentions of the network's actors.

The concept of shaping intentions is particularly of interest in order to achieve a given political or military goal. Influence campaigns seek to achieve political objectives through the conveyance of information and indicators with an intent to "influence the emotions, motives, objective reasoning, and ultimately the behavior [of others]"[DOD, 2003, pg. ix]. Focusing upon the important actor(s) and attempting to influence their behavior within a given environmental and situational context proves useful in a variety of applications including social, corporate and governmental (including non-military) endeavors [Renfro and Deckro]. Unfortunately, the most important actor, leader, or decision maker within a group is not always easily accessible, Osama bin Laden of Al Qaeda, for example.

This chapter continues the development of translating data observed and characterizing non-cooperative networks for use within mathematical programming applications, focusing primarily upon the concepts, measurement, and application of gains, losses, and thresholds of influence. These items serve as a continuation of the social network flow paradigm offered by Freeman et al. [1991], Renfro [2001], and Clark [2005]. Conceptual development of gains, losses, and thresholds of influence are discussed, and two potential means to measure this phenomenon are presented. A logical extension of the network flow based centrality measure offered by Freeman

et al. [1991] is explored. Finally, a demonstrative example illustrates the application of mathematical programming, while accounting for gains, losses, and thresholds of influence within a social network, to explore course of action analyses is demonstrated.

7.2 Social Network Flows

Influence campaigns seek to achieve political objectives through the conveyance of information and indicators to affect behavior. The social science literature is replete with descriptive theory capturing the nature and transfer of influence at an interpersonal level. Building upon these previous efforts, this chapter develops connections between social science's assessment of interpersonal communication and operations research's network flow formulation. Nuances of influence, specifically gains, losses, and thresholds, from a sociological perspective are described and then discussed within the context of a generalized network flow problem. Analysis of a notional social network is presented, including applications of classical sensitivity analysis to deal with uncertain data–an inevitability when dealing with clandestine organizations. The resulting methodology explores courses of action that seek to influence a potentially inaccessible target audience using their own indigenous social network as a conduit.

Throughout the field of sociology, the endeavor of identifying actor importance has been accomplished via a number of measures, relying primarily upon the application of graph theory to social networks [Wasserman and Faust, 1994, pg. 169]. The majority of these measures focus on characterizing actor importance using various properties of location with respect to all other actors within a network. The underlying assumption is that the most important, or most prominent, actors are usually strategically located within the network [Wasserman and Faust, 1994, pg. 169]. The values derived for these measures are often related to the flow or transfer

Table 7.1: SNA and Network Flow Relationships [Renfro and Deckro, 2003]

SNA Terms	Flow Model Properties
People	Nodes (sinks, sources, or transshipment)
Connectivity or affinity	Capacitated arcs between nodes
Social Closeness	Capacity
Influence	Commodity
Potential Influence	Magnitude of flow
Initiators of influence	Source(s)
Targets to be influenced	Sink(s)
Intermediaries involved	Transshipment node(s)

of information or the flow of influence among actors within a network [Freeman et al., 1991; Lopez et al., 2002; Renfro, 2001; Wasserman and Faust, 1994].

From the sociological perspective, the most important actors either control or have the ability to receive a greater amount of information relative to the other individuals within a social network. This chapter approaches actor importance from a slightly different perspective in that the most important actors will comprise the target audience of psychological operations. This could include an organization's decision makers or possibly a target population that is a subset of the overall network. The representation of influence as a transferrable commodity, however, carries over and serves as the basis of the methodology presented. Recall the mapping between SNA and OR, presented again in Table 7.1.

Renfro suggested that gains and losses of information or influence in the context of SNA are analogous to "...preconceived opinions or influence from outside the network being modeled ...predispositions of individuals favoring (or opposed to) the influence represented by the flow ...or communication problems such as misunderstanding the message" [Renfro, 2001, pg. 88]. These suppositions actually incorporate several areas within persuasion theory.

For example, "preconceived opinions" invokes the concepts of knowledge and reporting bias. The former "is the presumption that a communicator has a biased view of an issue" [Perloff, 2003, pg. 164]. Reporting bias is "the perception that

the communicator has opted not to report or disclose certain facts or points of view [Perloff, 2003, pg. 167]. Therefore, considering these types of biases, and minimizing their effects, is important when imposing an influence upon the access points to a target network.

The "predispositions" of individuals and the acceptance or denial of influence relates to the how successful the persuasion will tend to be. Perloff defines persuasion as "a symbolic process in which communicators try to convince other people to change their attitudes or behavior regarding an issue through the transmission of a message, in an atmosphere of free choice" [Perloff, 2003, pg. 9]. The reinforcement or change of a predisposition is, in effect, the influence. It is important to note that successfully influencing an individual could be a the result of persuasion or power, both of which contribute to the model development discussed in the next section.

Lastly, "communication problems" such as misunderstanding the message could simply be due to inter-media transfers (e.g., voice to transcript), inter-language transfers (e.g, translations and transliterations), or inter-personal communication media (e.g., errors or failures of communication devices or the users that implement them). This last area reinforces the fact that influence is a phenomenon that is specific to the actor and sender involved in the communication. Thus, the measurement of the forces that may modify the flow of influence within a social network is problematic. Even more complicating is the fact that a desired change in attitude or behavior is also dependent upon the setting within which the inter-personal exchange is made.

7.2.1 The Flow of Influence

The seminal work of French describes the formative theory of social power and analyzed and addressed some of its limitations. In the course of his work, French defined "the basis of interpersonal power... as the more or less enduring relationship between (two individuals) A and B which gives rise to power" [French, 1956, pg. 183]. He then described five bases for power: attraction, expert, reward, coercive,

and legitimate [French, 1956, pg. 183-4]. In examining the impact of peer group influence upon opinion formation, a now prevalent interpretation of French's work is that "[French] first proposed that social influence was a finite distributed resource" [Friedkin and Cook, 1990, pg. 130].

Friedkin and Cook discuss social influence in the context of interpersonal relations within a network and their subsequent role in the exchanges of influence required to enable opinion formation [Friedkin and Cook, 1990]. The resulting models essentially attempt to describe the personal interactions that transform a network of individuals with discrepant opinions into a network whose members' opinions have coalesced, at least to some degree. Similar concepts in social network literature that are based upon an exchange of influence between individuals include contagion (of behavior) [Leenders, 2002; Scherer and Cho, 2003], diffusion (the rate of acceptance of innovative and possibly risky ideas or behavior) [Valente, 1996], and the "infectious movement of desires and ideas from mind to mind" in the context of permission marketing [Buchanan, 2002, pg. 160-1].

Beginning with French's work, it is clear that the social science research and theory liken the interaction between two individuals and the resultant exchange of information, opinion, or influence to that of a commodity that flows between them. When utilizing social network analysis concepts such as tie strength (or social closeness) to serve as arc capacities in a flow of influence analysis, there are some instances where this measure alone may be insufficient to accurately determine the influence one individual has over the other. The next sections elaborate upon modeling such concepts–gains, losses and thresholds of social influence.

7.2.2 Gains and Losses

To frame the context of gains and losses of influence, it is critical to begin with a set of working definitions, as influence, persuasion, and power are frequently interchanged within the literature, despite their subtle differences. In a more recent

theoretical look at these concepts, Lovaglia et al. [2003] define *influence* "to occur when a person's opinion or behavior changes to conform to the suggestion of another without the threat of punishment or the promise of reward" [Lovaglia et al., 2003, pg. 109]. *Persuasion*, then, is a tool used to influence others. On the other hand, the concept of *power* implies the use of force, coercion, sanctions, or is derived from opposing interests and the availability of resources to those opposing parties [Lovaglia et al., 2003, pg. 109-10]. However, Goldhamer and Shils defines a powerful person as one who "influences the behavior of others in accordance with his own intentions" [Goldhamer and Shils, 1939, pg. 171]. Ultimately, applications of both power and persuasion seek to change or modify another's behavior or attitude; the difference between them lies in the method chosen to induce this change. Therefore, taking a more general approach, let *influence* be defined as follows.

Definition 12. *Influence: to induce a change in behavior of another that conforms to the influencing actor's desires, either cooperatively or otherwise*

Elements of both persuasion and power theory may now be used to explain, and serve as a basis for a mathematical model of, gains and losses of influence. The resulting measurements are associated with specific i-j pairs of actors. As expected, some individuals in a social network can be more (or less) influential, despite the strength, or closeness, of their relationships. Many situations may exist where influence is not necessarily equitable between the two actors, resulting in an interaction-dependent effect upon the information or influence exchange between two actors. For example, a father and son may have a strong bond between them, and therefore a relatively high value for tie strength. However, regarding the influence one has over the other, the father and son are likely unequal.

This approach differentiates this methodology from the one suggested by Clark [2005] regarding the effects of tie strength and influence differentials due to the sending actor's characteristics. Clark developed a combined measure of potential influence, with inputs comprised of topological and actor-specific effects of influence

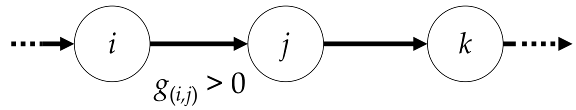


Figure 7.1: Arc with Multiplier

[Clark, 2005, pg. 3-34]. The method presented here contends that, as illustrated by the father-son example, tie strength between two individuals serves as an upper bound for the potential flow of influence between them–a true arc capacity. How easily an individual can promulgate influence to fill the arc to capacity is dependent upon the receiving actor's perception of power or persuasive ability inherent within the sending actor. Effectively, less (more) effort must be exerted by influential or powerful (less influential or powerful) people in order to promulgate the same amount of influence.

This suggests the information or influence flow through a network may require a multiplier to improve the requisite representation of network behavior in network flow models. This technique is borrowed from the generalized network flow formulation, which has been applied to physical systems exhibiting similar traits (e.g., spoilage of fruit during shipment, evaporation or collection of water during its movement through open irrigation canals, exchanges rates, and so forth).

Suppose that individual j has great respect for individual i such that individual i always has a tremendous impact (i.e., influence) upon individual j (e.g., actor i may be referred to as an *opinion leader*). This gain in influence on arc (i, j), illustrated in Figure 7.1, consequently requires that $g_{(i,j)} > 1$, and has the corresponding GNF constraint for node j: $x_{(j,k)} - g_{(i,j)}x_{(i,j)} \geq 0$. Subsequently, for every $x_{(i,j)}$ unit of flow pushed through arc (i, j), $g_{(i,j)}x_{(i,j)}$ units of flow arrive at node j.

Alternatively, suppose that individual j has never been impressed with individual i, possibly due to individual i demonstrating poor performance or untrustworthiness in past interactions; for example, the loyalty or honesty of actor i has been previously questioned. The corresponding GNF constraint for an influence loss on arc

(i, j) is identical to the gains constraint with the exception that $0 < g_{(i,j)} < 1$. Ultimately, the influence that actor i has over others is reduced from the level of influence that originally flowed out of individual i. The degradation could easily be attributed to personality conflicts, miscommunication, disrupted communication, and even the network structure itself [Friedkin and Johnsen, 2002; Lopez et al., 2002]. Degradation of flow through a network is also a common phenomenon observed within physical networks such as communications, energy, shipping, and irrigation [Ahuja et al., 1993, pg. 8].

Of course, there may exist any combination of the multipliers $g_{(i,j)}$ on the arcs within the given network. Potential means to place a number on a gain or loss of influence between two individuals could include evaluating individuals in the context of the bases of power, attraction, expert, reward, coercive, and legitimate, discussed in French [1956, pg. 183-4]. In addition, the work of Lopez et al. [2002], originally intended to explore the efficiencies, or more specifically the lack thereof, of information flow within the "traditional hierarchical topologies commonly used by organizations," developed an information dominance measure that could serve as a proxy for the gain multiplier [Lopez et al., 2002]. An approach, similar to the one taken by Clark [2005] to account for individual-specific effects of interpersonal influence, is described next.

7.2.3 Measurement of Gains

Since influence, as defined in Definition 12, essentially contains elements of power and/or persuasion theory, a model of influence gain should be able to account for either or both phenomenon. Such a concept exists: charisma. Charismatic leaders often share the fundamental communicator characteristics of authority, credibility, and social attractiveness [Perloff, 2003, pg. 152]. As seen in Figure 7.2, the characteristics contributing to charisma may involve elements of power, persuasion, or both. Mathematically modeling the charisma of a communicator is clearly prob-

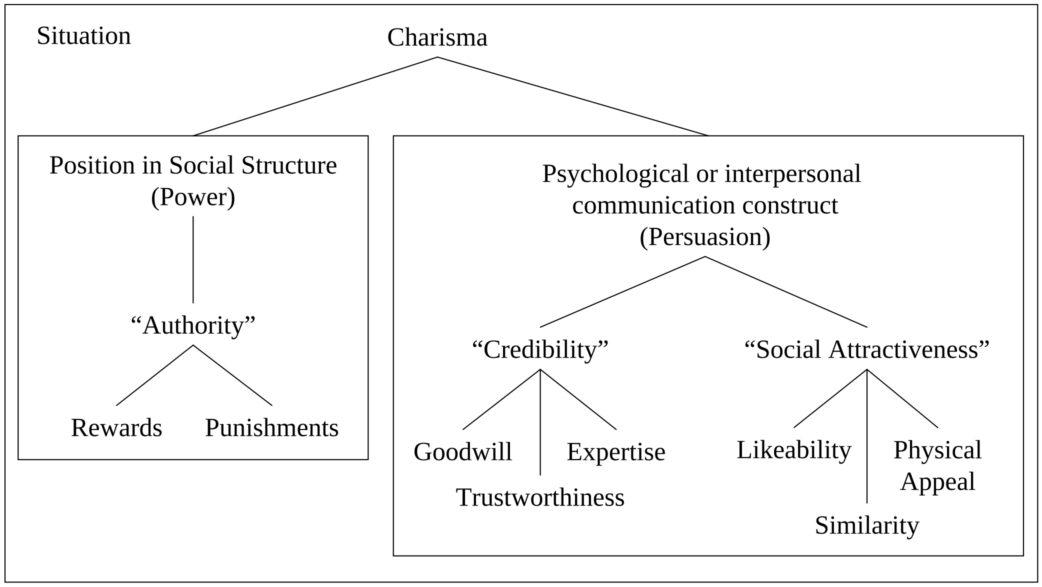


Figure 7.2: Charisma [Kelman, 1961; Perloff, 2003]

lematic, as not only are these concepts subjective, but the measurement and roles of each characteristic are dependent upon the situation within which the communication occurs [Perloff, 2003, pg. 159, 161-3]. In addition, "just as there is not one type of charismatic leader, there is not one defining characteristic of effective communicators" [Perloff, 2003, pg. 152]. The literature is replete with efforts attempting determine what factors or attributes significantly contribute to leadership, charisma, power, and persuasive ability [cf. Kelman, 1961; Anderson et al., 2001; Perloff, 2003, Chp. 6, among others]. The answer seems to be a resounding 'It depends'.

Nonetheless, a multiple logistic regression model offers an opportunity to measure the gain (or loss) between two individuals based upon their attributes that may contribute to, or serve as a proxy for, a certain qualitative level of charisma as illustrated in Figure 7.2. With regard to multiple logistic regression, "the response variable of interest has only two possible qualitative outcomes, and therefore can be represented by a binary indicator variable taking on values 0 and 1" [Neter et al., 1996, pg. 567]. For this model, let $Y_i = 1$ indicate that the ith individual is perceived as charismatic or a member of the organization's leadership structure, zero

otherwise. Further, let $\mathbf{X}$ denote p attributes hypothesized to contribute to charisma (or lack thereof); these predictor variables may be "quantitative or qualitative and represented by indicator variables" [Neter et al., 1996, pg. 581]. Examples could include time spent in a group, age, education levels, reputation, and so forth. With Equation 7.1, the multiple logistic regression model is given by Equation 7.2 [Neter et al., 1996, pg. 581-3].

$$\beta'\mathbf{X}_i = \beta_0 + \beta_1 X_{i,1} + \ldots + \beta_{p-1} X_{i,p-1} \tag{7.1}$$

$$E\{Y_i\} = \pi_i = [1 + \exp(-\beta'\mathbf{X}_i)]^{-1} \tag{7.2}$$

With $\mathbf{b}$ denoting the maximum likelihood estimate of β, the fitted logistic response values are

$$\hat{\pi}_i = [1 + \exp(-\mathbf{b}'\mathbf{X}_i)]^{-1}. \tag{7.3}$$

The estimated logistic response function, or fitted value, for actor i, denoted $\hat{\pi}_i$, is interpreted as the estimated probability that actor i, with the given characteristics $\mathbf{X}_i$ is charismatic, a leader, or to whatever the a priori definition of $Y_i = 1$ has been set [Neter et al., 1996, pg. 577]. The probabilities $\hat{\pi}_i$ serve as inputs to

$$g_{ij} = 1 + \hat{\pi}_i - \hat{\pi}_j. \tag{7.4}$$

When the transmission of influence is being sent from a less charismatic individual i (low $\hat{\pi}_i$) to a higher or more likely charismatic individual j (higher $\hat{\pi}_j$) then the gain multiplier tends to be less than 1, indicating a loss or degradation of influence. Ultimately, given an identical amount of influence required to pass between two individuals, less influential (or charismatic) individuals must exert more energy than those who are more charismatic.

Since $\hat{\pi}_i \in [0,1]$, $g_{ij} \in [0,2]$. Recall that $g_{ij} = 1$ implies neither a loss nor a gain of the commodity as it flows along the arc. A multiplier of $g_{ij} < 1$ implies

		Individual j	
		Charismatic	$\overline{\text{Charismatic}}$
Individual i	Charismatic	$\hat{\pi}_i \approx \hat{\pi}_j \Rightarrow g_{ij} \approx 1$	$\hat{\pi}_i > \hat{\pi}_j \Rightarrow g_{ij} \in (1,2]$
	$\overline{\text{Charismatic}}$	$\hat{\pi}_i < \hat{\pi}_j \Rightarrow g_{ij} \in [0,1)$	$\hat{\pi}_i \approx \hat{\pi}_j \Rightarrow g_{ij} \approx 1$

Figure 7.3: Gain Domain Based on $\hat{\pi}_i$

a loss, and $g_{ij} > 1$ implies a gain. Suppose $Y_i = 1$ if individual i is designated to be charismatic–and therefore influential–or 0 otherwise. If two actors i and j are peers with similar traits or characteristics that contribute to charisma, then it is assumed that they will have an equitable exchange of influence, or $g_{ij} \approx 1$. Note that this also applies to peers that are both non-charismatic. Consequently, small differences between individuals having similar charismatic or influential scores $\hat{\pi}_i$ and $\hat{\pi}_j$ will tend toward a gain multiplier of 1. Values of g_{ij} between peers not equal to one is to be expected, as Shamir noted that several theories regarding charismatic leadership "share the assumption that such leadership can be found at all levels of the organization" [Shamir, 1995, pg. 20]. The different possible scenarios are summarized in Figure 7.3.

The special case where $\hat{\pi}_i = 0$ and $\hat{\pi}_j = 1$, resulting in $g_{ij} = 0$ could be the mathematical equivalent of "Chicken Little," who is so uninfluential, so unpersuasive, so uncharismatic that the receiver j will ignore the influence emanated from actor i. It is also important to recognize the upper bound is 2 when applying this model. If such a multiplier is deemed insufficient (in magnitude), subject matter expertise will be required to differentiate these cases from those determined by Equation 7.4. Additionally, the specific components of Equation 7.1 are unlikely to be completely determined a priori; therefore, the definitive list of traits that mathematically characterize an influential individual are more likely to be predicated upon the, presumably

limited, data available. Fortunately, techniques exist to test hypotheses regarding the statistical significance of such traits, means to deal with missing data, and other methods to gain new insight into the data and the organization [cf. Little and Rubin, 1987].

Finally, the incorporation of gains into a generalized network flow model of a social network is a logical and relatively straightforward extension of the centrality measure developed by Freeman et al. [1991] (which is also similar to the one developed by Brandes and Fleischer [2005]). The measure developed by Freeman et al. evaluated various maximum flow characteristics of the network using hypothesized arc values representing the strength of a relationship. These arc capacities were simply assumed to be available. With the measures of tie strength developed in Chapter VI supplementing the previous methods developed by Renfro [2001] and Clark [2005] along with the inclusion of gains and losses due to persuasion presented in this chapter, a generalized and assumed more representative network flow model may be analyzed. MATLAB functions to perform network flow and generalized network flow centrality are provided in Appendices K and L, respectively. Both measures are applied to the data set discussed in Chapter VIII. The next section explores the use of right-hand side values in GNF to model influence thresholds.

7.2.4 Thresholds

Numerous social science researchers allude to the existence of what could be referred to as a *threshold*–a point at which an individual decides between two or more competing alternatives that is dependent upon both external influences and internal principles [cf., Buchanan, 2002, pg. 158]. The right-hand side values (b_i) can be used to model thresholds.

When the right-hand side coefficient (b_i) is less than 0 it signifies that node i seeks a demand of the commodity, in this case influence, of the absolute value of that amount. This notion of demand is one way to model thresholds of influence required

for an individual to pass, or fail to pass, influence further through the network. Again, the bases of attraction, or lack thereof, proposed by French [1956] may serve as a foundation for threshold measurement. Social position (e.g., a gatekeeper) may also serve as an indicator that some threshold of influence, passed from one or more other individuals, exists. Freeman describes the gatekeeper as one who is "not conceived as being in a general sort of position of control like a position high in centrality based on betweenness. Instead, he or she is the keeper of the gate controlling communication to and from a particular other person vis-à-vis the rest of the network" [Freeman, 1980, pg. 586].

Another view of thresholds is due to Granovetter, who suggests that "an actor has two distinct and mutually exclusive behavioral alternatives" [Granovetter, 1978, pg. 1422]. The alternative ultimately chosen is dependent upon the costs and benefits of that choice, the values of which are derived by the number of others observed making the same choice [Granovetter, 1978, pg. 1422]. Using the predilection to join a riot as an example, Buchanan observed that "The level of someone's threshold (to join a riot) would depend on their personality, and on how seriously they take threats of punishment, for example" [Buchanan, 2002, pg. 107] [cf., Granovetter and Soong, 1983, pg. 166].

In the context of this methodology, once the threshold is met, that is, once a certain amount of external influence has flowed to, and is absorbed by, that individual from his or her networked peers, that individual will accept the influence as beneficial enough to propagate further. All of these aspects essentially comprise the data elements required for model input: a value for the threshold, the potential interactions, and the levels of influence given an interaction has occurred.

A number of interesting network behaviors can be captured via thresholds. Consider the GNF analogy, where actor j requires some amount of influence ($b_j < 0$) before passing influence, relaying a message, or exhibiting an influential behavior to

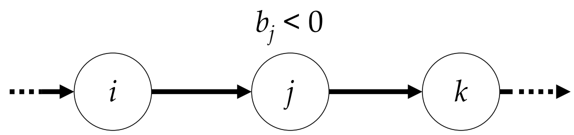


Figure 7.4: Node Demand (Threshold)

actor k. The general concept is illustrated in Figure 7.4; the corresponding constraint is $x_{(j,k)} - x_{(i,j)} \geq b_j$.

An extreme variation of this case may include an *absorbing* node. For such a node, the demand for influence must be at least as great as the sum of the capacities of all arcs entering that node. For a given node j, let the absorption value U_j be

$$U_j = \sum_{j:(i,j) \in A} u_{(i,j)}. \tag{7.5}$$

From Figure 7.4, if $b_j = -U_j$, the individual j would prevent influence from passing on to actor k. The mathematical formulation for this constraint in the GNF for any given node j is $x_{(j,k)} - x_{(i,j)} \geq -U_j$. Such an individual may be likened to an overzealous gatekeeper or someone who is isolating actor k from other influences.

Demand at any given node i (b_i) is also analogous to the threshold of an individual's vulnerability to be influenced by others within the network. For example, some individuals, due to position within the organization, psychosocial tendency, and so forth, will readily accept influence and immediately promulgate it to others with whom they are connected. Essentially, these individuals simply serve as a pass-through where $b_i = 0$, also known as a transshipment node. The GNF formulation for this case is simply the conservation of flow, $x_{(j,k)} - x_{(i,j)} = 0$. Most individuals are unlikely to behave as an overzealous gatekeeper; others are more savvy regarding the likelihood of an influence campaign against their organization. Therefore, the most likely range for an individuals threshold is $-U_j < b_j \leq 0$.

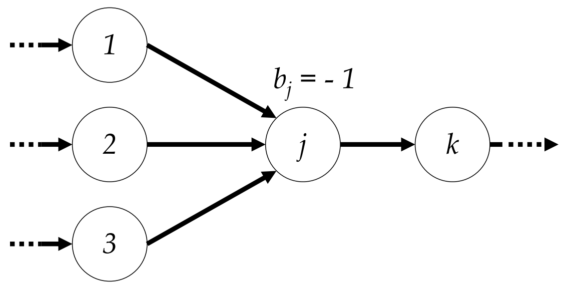


Figure 7.5: Conditional Gatekeeper

An alternative approach to thresholds that permits a m-out-of-n reporting scheme may also be accomplished via the use of right-hand-side values. For example, suppose that an individual must receive m independent reports out of n possible sources. In this setting, "reports" could comprise a flow of influence via a message. Based upon organizational procedures, individual j uses independent sources to confirm the authenticity or importance of the influence prior to forwarding the message deeper into the organization. The actor essentially serves as a conditional gatekeeper; influence is passed on if a specified condition is met. This approach is illustrated in Figure 7.5. Suppose that no gains or losses are evident in the arcs between 1, 2, 3, j, and k. Further, assume that flow on these arcs is either 0 or 1. The network structure in Figure 7.5 essentially depicts the m-out-of-n situation. The first unit of influence sent to actor j is consumed by the demand $b_j = -1$. Therefore, actor j must get information or influence from at least 2 of the 3 individuals before information or influence is passed on to actor k.

7.2.5 Costs

The final variable within the GNF formulation to review is cost per unit flow along the arcs $c_{(i,j)}$. Developing the cost per unit flow along the arcs requires a local perspective from each actor. If influence is to pass between two individuals (i, j), then actor i must internally consider the various costs and benefits of transmitting the message or influence, and actor j must internally decide his or her own costs and

benefits of forwarding the influence further. These concepts are reflected in social action theory.

In studying the impact of social network structures upon their internal information flow, Yamaguchi [1994] emphasizes the importance of social action in the diffusion of information within a social network. He states two main reasons for this concept's relevance from both a receiving and transmitting viewpoint. When receiving information or influence, "actors evaluate information and act according to the results of their evaluations" [Yamaguchi, 1994, pg. 59]. This evaluation may include consideration of who transmitted the information, in what fashion, in what context, at what level of emphasis, or any number of situation-dependent perspectives. When transmitting influence, a more relevant concept when trying to ascertain the potential costs for the GNF model, the action "will depend not simply on the presence of communication between them, but on the rational assessment of costs and benefits regarding the exchange of information [Yamaguchi, 1994, pg. 59].

Therefore, the mere existence of a communication path does not necessarily guarantee that information or influence will flow freely. Although the specific concept of 'costs and benefits' was not further elaborated upon by Yamaguchi [1994], this concept appears to accommodate important aspects regarding costs in the GNF formulation. Such costs would include each actor's own assessment of the risks associated with enacting a communique, the actual cost to transmit the influence (e.g., email, long-distance call, satellite call, travel via donkey through a treacherous mountain pass, and so forth), and other factors (e.g., the possible 'penalty' for passing on rumors or propaganda in a given environment) as necessary. The costs stemming from a perspective external to the network are similar in nature, but are assessed as those costs as perceived by the organization trying to pass influence into the network. These may also incorporate likelihood of mission success from either perspective.

Measurement of influence in the context of SNA is "based upon the importance of relationships among interacting (individuals)" [Wasserman and Faust, 1994, pg.

4]. One of the underlying principles of SNA is that "...individuals view the network structural environment as providing opportunities for or constraints on individual action" [Wasserman and Faust, 1994, pg. 4]. This implies that when an individual within a network comes upon a decision point, they tend to take certain individual's opinions (e.g., those socially close) or authority into account. There are a variety of examples in SNA literature that investigate and attempt to measure this influence [cf., Frank and Yasumoto, 1988; Friedkin and Cook, 1990, among others].

7.2.6 Solution Procedures

When studying social networks as generalized network flow problems there are a range of choices of solution approaches. First, the problem may be formulated and solved as a linear program (LP). Alternatively, the problem may be solved via the network simplex algorithm. Although the two approaches share the same underlying requirements for feasibility and optimality, the latter is "200-300 times faster" [Bazaraa et al., 1990, pg. 419]. Due to relatively simple nature of the notional example in the next section, the analysis is conducted via a LP formulation and solution. Use of a LP approach facilitates conducting post-optimality analyses that provide a means to deal with the underlying uncertainty in input data.

7.2.7 Network Flow

Two mathematical formulations of the network flow problem, both characterizing influence as the commodity, are of interest: the maximum flow (MF) and the generalized network flow (GNF) problems. MF, a special case of GNF, determines the maximum flow that can pass from one or more source nodes to one or more sink nodes. MF is useful in determining whether or not any flow between two individuals or groups of interest through the network is feasible. MF has also been used as a basis for a social network analysis measure using valued social networks as an input [Freeman et al., 1991]. The work of Freeman et al. [1991] represents one of the early

departures in SNA from binary representations of relationships (they either do or do not exist) to one that accounts for the strength of interpersonal relationships when trying to estimate actor importance. For their SNA measure, Freeman et al. [1991] used a value representing the strength of a relationship as an arc capacity in a social network; they then applied MF to each actor, using the actor as a source and all other actors as sinks, in order to ascertain actor importance. While Freeman et al. [1991] simply assumed these values would be made available, Renfro [2001] devised a means to estimate the strength of a relationship and then applied various approaches, including MF, to gain insight into clandestine social networks. An additional application of MF to clandestine social networks is found in Clark [2005], who combined a topology-based SNA measure with information derived from individual characteristics to estimate relationship strengths.

The mathematical model of primary interest in this chapter is GNF, which satisfies a predetermined amount of flow between one or more source nodes to one or more sink nodes in an oftentimes least-cost manner. The GNF provides a means to model a variety of real-world networks with commodities that undergo degradation or improvement over time or distance. In the context of interpersonal communication, one such degradation process could include the content and context of a rumor spread throughout a social network, where the message received by the second person could be significantly garbled when received by the twentieth person in a chain or path. The inherently flexible nature of GNF, particularly the ability to model changes in commodity levels during its travel through a network, lends itself to capturing the phenomena of gains, losses, and thresholds of influence within a social network.

Building on the work of [Freeman et al., 1991; Renfro, 2001; Renfro and Deckro], this research continues the development of the parallels between the flow of commodities in the physical world and the flow of influence in the behavioral realm. To begin, the generalized network flow problem formulation first presented in Section 2.6.3 is repeated [Ahuja et al., 1993, pg. 567-8].

Table 7.2: GFP Variable Definition

Variable		Definition
$c_{(i,j)}$	$\equiv$	the cost per unit flow induced from node i to node j
$x_{(i,j)}$	$\equiv$	number of units of flow from node i to node j on arc (i,j), $x_{(i,j)} \in \left[0, u_{(i,j)}\right]$
b_i	$\equiv$	0 if node i is a transshipment, or 'pass-through,' node; < 0 if demand is required by node i; and, > 0 if supply is provided from node i
$g_{(i,j)}$	$\equiv$	a rational value $> (<) 1$ that indicates if arc (i,j) is gainy (lossy); if $g_{(i,j)} = 1$, then the arc (i,j) is neither one
N	$\equiv$	the set of nodes (individuals) within the network
A	$\equiv$	the set of arcs (i,j) (connections between individuals) that form the network

$$Minimize \quad \sum_{(i,j)\in A} c_{(i,j)} x_{(i,j)} \tag{7.6}$$

$$\sum_{\{j:(i,j)\in A\}} x_{(i,j)} \quad - \sum_{\{j:(j,i)\in A\}} g_{(j,i)} x_{(j,i)} \quad \geq \quad b_i \qquad \forall \quad i \in N \tag{7.7}$$

$$0 \leq x_{(i,j)} \leq u_{(i,j)} \quad \forall \quad (i,j) \in A \tag{7.8}$$

The objective function, Equation 7.6, seeks to minimize the total cost of flow through the network, subject to the mass balance and arc capacity constraints– Equations 7.7 and 7.8, respectively. Note that the constraints 7.7, which replace the equality (=) with (≥) from the traditional mass balance constraint, allows for potential violations of traditional conservation of flow assumptions. This extension, a relaxation of the original formulation, facilitates feasibility, particularly when gains and losses affect flow (e.g., 1 unit enters and, due to gains, say 2 or more must exit) and when arcs are capacitated (e.g., there exists a maximum amount of flow that may traverse the arc, social closeness serving as an upper bound in this case).

7.2.8 Underlying Assumptions

This approach equates influence to the commodity that flows through the network; the amount of influence traveling from actor i to actor j is denoted by $x_{(i,j)}$. The greater the magnitude of $x_{(i,j)}$, the greater the relative influence exerted upon individual j by individual i.

In addition, social closeness is defined as a positive, real-valued estimate of the "maximum potential influence one person i has upon another person j in a set of N people in a given scenario;" this serves as the capacities of potential influence, denoted $u_{(i,j)}$ in the GNF formulation [Renfro, 2001, pg. 89].

The remaining parameters, $g_{(i,j)}$, b_i, and $c_{(i,j)}$ are the focus of this chapter. The arc multiplier $g_{(i,j)}$ provides a means to model gains and losses of influence. The demand variables b_i, when associated with a transhipment node (i.e., an individual that is neither a source nor a sink) may be used to model individual thresholds. Relationships between the cost coefficient $c_{(i,j)}$ and operational risks associated with interpersonal communication are suggested. Such costs could represent one of two types of risk: either internally among the organization's members, or externally as an aspect of initiating or determining the efficacy of an influence operation course of action, for example. Other costs are possible as warranted.

A course of action is defined as a psychological operation that attempts to influence an accessible subset of actors within a social network of interest in order to influence the behavior of any number of actors. The targeted individual(s) are perhaps the most important or respected leader(s), or a disgruntled element that may not be directly accessible. The course of action essentially determines which accessible actors serve as conduits (sources) and which oftentimes inaccessible actors serve as targets (sinks). The sources facilitate the insertion of an external influence into an adversarial social network. The sinks comprise the ultimate targets of influence; they could be senior leaders, decision makers, specific subgroups, or anyone who has the ability to affect the overall behavior, actions, and objectives of the entire

network. All other actors that may be used to promulgate the influence from the sources to the sinks are considered as transshipment nodes. The overall goal is to influence the target population such that they alter their behavior as desired.

Obviously, obtaining accurate and complete intelligence detailing such nuances of a non-cooperative (and likely covert) network is a formidable task, especially when the nature of the adversary as well as the nature of compartmentalized intelligence agencies is considered. Examples of these challenges are (painfully) described in detail in the recent *9/11 Commission Report* [National Commission on Terrorist Attacks Upon the United States, 2004, pg. 71-102]. Over a decade earlier, Sparrow explored "the opportunities for the application of (social) network analytic techniques to the problems of criminal intelligence analysis, paying particular attention to the identification of (organizational) vulnerabilities..." [Sparrow, 1991, pg. 251]. He conceded that missing data (for any number of reasons), fuzzy and ambiguous boundaries of inclusion or exclusion of individuals, and the inherently dynamic nature of human interaction (and therefore social network composition) add a high level of complexity to this problem [Sparrow, 1991, pg. 261-2]. Related works investigating the impact of missing information have concluded mixed, but oftentimes detrimental, effects upon the analyses [Bolland, 1988; Borgatti et al., 2006; Costenbader and Valente, 2003; Sterling, 2004; Thomason et al., 2004].

Despite these findings, this research assumes that the measures, or at least estimates of the measures, involved in characterizing networks of interest are indeed obtainable due to the increasing interest in, and consequential approaches to, this problem as suggested by Dombroski and Carley [2002]. An advantage to applying the network flow models from operations research to those characterizing social networks is that a number of techniques are available to evaluate the sensitivity of essentially all model inputs. As in traditional mathematical programming applications, this offers a means to account for some of the uncertainty likely to be found in these sociological measurements.

7.3 Notional Example

The overarching objective of this type of analysis seeks the development and estimating the efficacy of various courses of action that attempt to influence accessible individuals in order to indirectly influence potentially inaccessible decision makers. Advantages inherent within the linear programming approach, the ability to perform post-optimality analyses for example, facilitate the investigation of uncertainty and its effects upon the results.

The situations and parameters discussed in this example are entirely notional and are only for demonstrative purposes. Assume that information has been gathered on a network of 11 individuals, with the communication or interaction between them indicated by the arcs (i, j). For this example it is assumed without loss of generality that the information will be one-way, with the direction indicated by the arrow. The flow of influence or information along a given arc (i, j) is bounded by $0 \leq x_{(i,j)} \leq u_{(i,j)}$, where $u_{(i,j)}$ are determined by an assessment of social closeness. All costs incurred per unit of influence flow along an arc are also assumed available.

The network of interest is illustrated in Figure 7.6. For each arc, the capacity upper bound and cost are denoted by $(u_{(i,j)}, c_{(i,j)})$, respectively, and are shown near their corresponding arc. Gains and losses are indicated by values within a triangle adjacent to the applicable arc. The only threshold modeled in this example is the one for individual 4 ($b_4 = -1$). Further assume that direction of communication is known, which is indicated by the directed arrows.

The next step is to determine sources and sinks for the flow of influence through this network. As opposed to the evolution of attitudes through the dyadic interaction of individuals over time (the focus of the majority of social sciences network models), the intention in this illustration is to force an influence through the network. This approach is well suited to gaining a better understanding of how an organization, or a subset of its individuals (e.g., the leadership), could change its attitudes in a

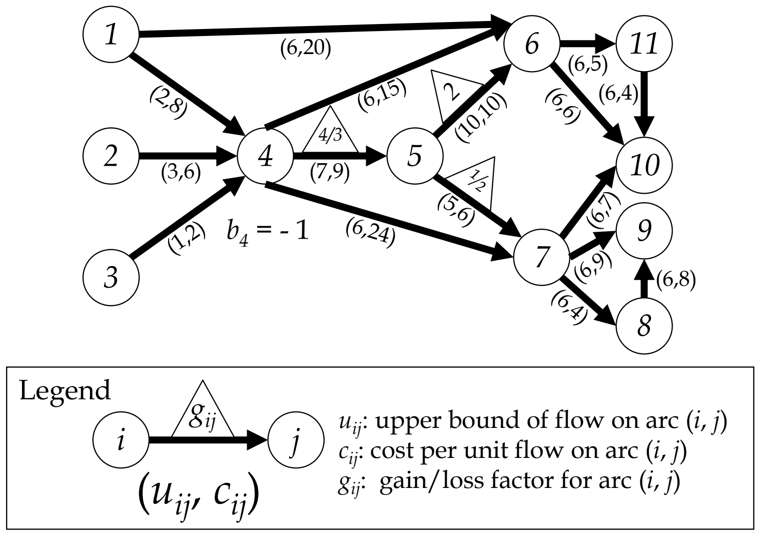


Figure 7.6: Notional Social Network

manner consistent with an external organization's interests, in essence, political force or influence. The ultimate goal, in this example, is to influence a number of target individuals to at least a minimum level and, through modeling, understand what actions may be required to do so in terms of cost and influence campaign activities. If the target individuals comprise the network leadership, they may not be directly accessible and therefore not immediately vulnerable to these operations. However, subordinates that report to these leaders may not only be vulnerable, but more easily accessible, and (much more importantly) trusted by those already within the network. Therefore, the overall strategy is to influence the more vulnerable (and accessible) actors in such a manner that the target individuals eventually receive the message (influence) in such a way that their opinions are changed to meet the overall political goals intended by the initiator of the influence campaign.

Assume that an initial assessment of the organization revealed that the set of individuals $\{9, 10\}$ were the leaders or the most influential members (opinion leaders) of the group having the ability to influence in this context. Influencing them would ultimately provide favorable results. However, by the very nature of their positions, they are shielded from such actions by individuals outside their own

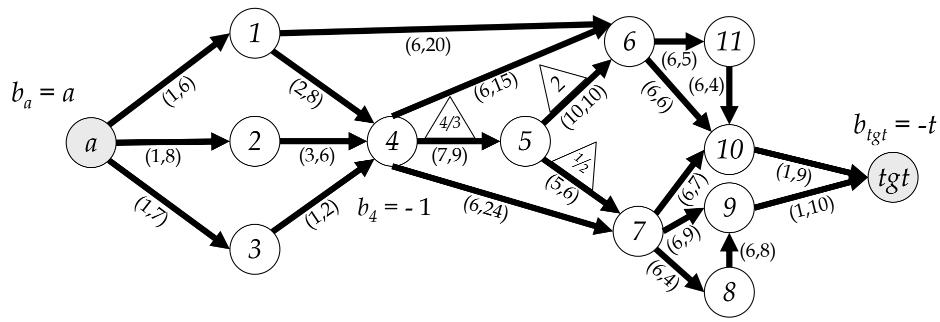


Figure 7.7: Notional Network with Target Sets

social network. Further, assume that the set of individuals $\{1, 2, 3\}$ are accessible and there exists a possible path of influence between them and the desired target audience. Therefore, a course of action (a) is devised to (1) identify accessible individuals that can be influenced and the the concomitant operational risks involved in doing so; (2) estimate the amount of influence to be pushed to these individuals ($b_a = a$); (3) identify the inaccessible individuals that comprise the ultimate target set (tgt); and, (4) develop a means to determine the efficacy of the operation through observation of target behavior ($b_{tgt} = -t$). This results in the network representation shown in Figure 7.7.

A variety of issues may be explored via this modeling and analysis approach. For example, course of action (a) is not necessarily limited to using the set of individuals $\{1, 2, 3\}$ to initiate an influence operation, but merely those that are accessible by another party external to the social network of interest. The amount of influence for this course of action is currently capacitated at unity for each arc emanating from node a. Again, this is not a general requirement. Two or more messages or attempts to influence an individual (e.g., repeated threats, emails, reminders, etc.) may be required or desired; this may correspond to a capacity of 2 or more. Similar arguments may be made for the individuals that comprise the target audience. Note that the super sink node (tgt) is a means to assess the overall effectiveness of the course of action. For this example, flow into this node (tgt) implies that at least one

of the actors $\{9, 10\}$ has demonstrated (verbally, physically, politically or otherwise) that they have received and responded in some manner to the influence initiated by the course of action.

Formulations that are infeasible directly translate to undesirable courses of action for any number of reasons; these might include insufficient external influence to obtain desired overall effects, insufficient connectivity within the network, poor choice of vulnerable nodes, unrealistic or unobtainable levels of effect required on the target audience, and so forth. Maximum flow formulations provide a means to verify the potential success of a given course of action. Note that due to the possibility of thresholds and multiple sources or sinks, the classic maximum flow formulation no longer applies. However, such problem aspects have been addressed in the maximum flow literature. For example, Megiddo [1974] generalized the maximum flow problem to optimal and fairly distributed flow among multiple sources and sinks; Miller and Naor [1995] developed a maximum flow algorithm for a network with known supplies and demands. With this in mind, references to a maximum flow formulation in this chapter are assumed to accommodate such extensions.

Given the chosen sets a and tgt and any appropriate thresholds, if the maximum flow is infeasible then the GNF formulation of the same network will also be infeasible. Using the same network shown in Figure 7.7 while focusing on the connectivity aspects of the target network results in the network shown in Figure 7.8; the upper bounds for arc capacities are denoted as $[u_{(i,j)}]$ and v denotes the value of flow to be maximized.

The optimal solution for the maximum flow problem of Figure 7.8 is $v = 3$; the resulting flow is shown in Figure 7.9, where the bold arrows indicate arcs with flow of 1, zero otherwise. This result establishes the feasibility of the current course of action, showing that it is possible to pass influence to both nodes 9 and 10, but it may require the initiation of influence through all three action nodes–1, 2, and 3.

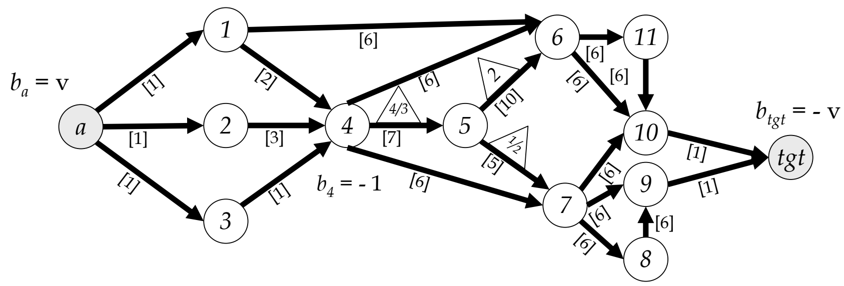


Figure 7.8: Notional Network (Maximum Flow)

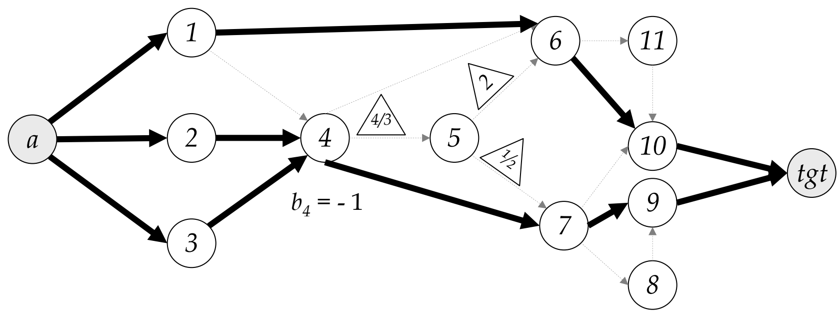


Figure 7.9: Maximum Flow Solution

Using the information from the maximum flow results, the incorporation of costs is likely to provide a more realistic flow pattern [Yamaguchi, 1994]. The solution to the minimum cost maximum flow problem of our notional network using $b_a = 3$ and $b_{tgt} = -2$, has an objective value (total cost incurred due to flow of influence through the network) of 93.33 units. The solution is shown in Figure 7.10, where highlighted arcs indicate flow; the specific amounts of flow are shown in braces, $\{x_{(i,j)}\}$, alongside the respective arc (i, j).

As suspected, the flow through the network accounting for costs is different from that of the maximum flow formulation. Note that conservation of flow is maintained for both the threshold and gains and losses effects. Interestingly, contacting, or influencing, all three susceptible nodes $\{1, 2, 3\}$ does not fully meet the target node objectives (i.e., $\frac{4}{3} + b_{tgt} = \frac{4}{3} - 2 < 0$). This is primarily due to the relaxation of the mass balance constraint. An interpretation of this is that if the network operated

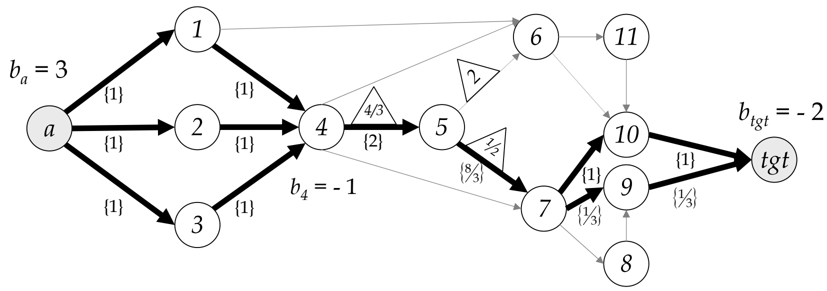


Figure 7.10: Minimum Cost Maximum Flow Solution

in a minimum cost manner, the current course of action to influence all susceptible actors may not result in meeting the overall objectives. Changing the constraint for the *tgt* node from '>' to '=' would remedy this effect but at a greater cost, assuming the problem remained feasible.

The solution basis plays a major role in post-optimality analyses. A change in basis often results in a different objective function value, except in the case of multiple optimal solutions. Degeneracy, indicated by a basic variable with a value of 0, is a common occurrence in network problems due to the balance constraints, particularly when $b_i = 0$ [Gal, 1979, pg. 314]. This situation is observed in this example, with arcs $(6, 11)$ and $(7, 8)$ in the basis, yet $x_{(6,11)} = x_{(7,8)} = 0$. Under these conditions, caution must be used in interpreting the shadow prices–an aspect of the information available for post-optimality analyses [Bazaraa et al., 1990, pg. 258] and [Martin, 1999, pg. 99].

Another important aspect of the current solution to address is the non-tree arcs at capacity. For example, flow along arc $(3, 4)$ is at capacity, yet $x_{(3,4)}$ is not in the basis. Consequently, changes in the arc capacities as part of the post-optimality analysis must first discern whether it is a tree- or non-tree arc, that is, one must identify whether the arc is or is not in the current basis, respectively.

7.4 Post-Optimality Analysis

Post-optimality analysis allows the investigation of changes in input data and is useful when uncertainty exists in the input parameters. Considering the intrinsic nature of the input data characterizing clandestine social networks, answers to some of the "what if" questions regarding gains, losses, thresholds, risks (both internal and external to the network members), estimates of the strengths of interpersonal relationships (social closeness), potential courses of action, and the impact of network connectivity, it is desirable to investigate the consequences due to changes among these parameters via post-optimality analysis.

Ahuja et al. highlight that there are primarily two approaches to network post-optimality analysis: combinatorial methods that re-solve a number of problems, and simplex-based methods that exploit information resulting from the linear programming algorithms [Ahuja et al., 1993, pg. 337]. Since the solution procedure chosen is LP-based, the simplex-based method is applied to the notional network solution.

Specific examples of consequences due a change in inputs may include no change to the current solution, a change in the objective value, a change in the basis (i.e., the influence may potentially take a different path through the network), a combination of change in objective value and change of basis, or overall infeasibility. It is also important to note that unless a parametric programming approach is specified, any allowable ranges are assumed to be applicable only as one-at-a-time variations to the original problem.

7.4.1 Changes in Gains and Losses

Unfortunately, post-optimality analysis of technological coefficients–the values that represent gains and losses of influence in this application–is not automatically provided by many optimization software packages. However, Bazaraa et al. provide a straight-forward approach to evaluating the excursions of a change to a single

column vector $\mathbf{a}_i$, which is applicable to both basic and non-basic variables [Bazaraa et al., 1990, pg. 282-3]. Nonetheless, their approach assumes that the new value is known *a priori*. Since that is likely not the case when dealing with uncertainty in intelligence data, analysts may take advantage of the parametric approach described in Gal [1979, Chp. 8] to ascertain the range of values that would satisfy the condition of interest, in this case a change of basis.

As observed in Figure 7.10, arc $(1,6)$ may be interpreted as a work-around for the gatekeeper, individual 4. Suppose it is of interest to investigate the implications of a gain for arc $(1,6)$ such that $g_{(1,6)} = 1 + g$, where $g > 0$ implies the current arc from actor 1 to actor 6 becomes *gainy* and $g < 0$ implies the arc becomes *lossy*. From this, the new column vector $\mathbf{a}'_{(1,6)}$ is formed:

$$\mathbf{a}'_{(1,6)} = [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad -(1+g) \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0].$$

To find the revised column vector $\mathbf{y}'_{(1,6)}$ given the current basis inverse $\mathbf{B}^{-1}$, the formula $\mathbf{y}'_{(1,6)} = \mathbf{B}^{-1}\mathbf{a}'_{(1,6)}$, is applied, yielding $\mathbf{y}_{(1,6)} =$

$$[\,0 \quad 0 \quad 1 \quad 0 \quad 0 \quad (1+g) \quad -(\tfrac{1}{3}+g) \quad -(1+g) \quad 1 \quad 0 \quad \tfrac{4}{3} \quad -(\tfrac{1}{3}+g) \quad -(\tfrac{1}{3}+g)].$$

Now check to see if this change would result in $x_{(1,6)}$ entering the basis (i.e., a different solution is required). Define $\mathbf{c_B}$ as the cost coefficients of the current basic variables. The non-basic variable $x_{(1,6)}$ must enter the basis if the condition

$$\mathbf{c_B}\mathbf{B}^{-1}\mathbf{a}'_{(1,6)} - c_{(1,6)} = \mathbf{c_B}\mathbf{y}'_{(1,6)} - c_{(1,6)} > 0 \tag{7.9}$$

holds.

The calculations result in $g < -0.0185$. Therefore, if even the slightest loss is imposed upon the arc $(1,6)$, the variable $x_{(1,6)}$ enters the basis, and a new optimal solution results, all other values remaining constant. As an example, suppose $g_{(1,6)} = 0.97$; the new objective function value is 93.075, $x_{(1,6)}$ enters the basis, and the total flow to the *tgt* node is improved to 1.64. The improved objective function is essentially due to cost savings resulting from a diminished amount of flow along

arc $(6, 10)$, as well as a more direct path to target actor 10. The improved amount of flow is due primarily to the fact that not all of the flow must pass through the gatekeeper (actor 4) and then incur the loss between actors 5 and 7. Note also that imposing a gain on this arc such that $g_{(1,6)} = 1 + g$, $(g > -0.0185)$ will not result in arc $(1, 6)$ entering the basis, *ceteris paribus*. Therefore, this suggests that a change in the underlying model assumptions, and their associated parameters, may assist in achieving the overall mission.

Using this information and technique, ranges may be developed for all columns, or variables, of interest. Unfortunately, if the problems are of reasonable (i.e., practical) size, this type of analysis may actually lend itself to re-solving a modified problem rather than implementing this approach, particularly if the ranges suggest further exploration and a new solution is desired. Methods dealing with other conditions, such as non-binding constraints, to determine one-at-a-time sensitivity ranges for the technological coefficients are described in detail within [Hartley, 1976; Bazaraa et al., 1990, pg. 281-3].

7.4.2 Changes in Thresholds

The right-hand sides (RHS) represented by b_i capture one of three phenomena: thresholds, sources, and sinks. The caution underlying this aspect of post-optimality analysis, in the presence of primal degeneracy, is the possibility that a right-hand side value may actually be increased or decreased beyond the range reported by optimization software and still maintain the current basis as optimal [cf., Sounderpandian, 2001]. Although degeneracy may adversely affect the sensitivity analysis in this manner, optimization packages typically mitigate this by simply providing a conservative window of allowable ranges. For the notional problem, these conservative RHS ranges are presented in Table 7.3. For node a the amount of influence pushed to the network can be reduced by at most 0.5 units before the current basis may change. Additionally, the amount of influence provided at node a cannot

Table 7.3: RHS Analysis

Row (Node)	Current Value	Allowable Increase	Allowable Decrease
a	3	0	0.5
1	0	1	0.5
2	0	1	0.5
3	0	0	0.5
4	-1	1	0.5
5	0	1.33	0.67
6	0	0	0.5
7	0	0.67	0.33
8	0	0	0.33
9	0	0.67	0.33
10	0	0.67	0.33
11	0	0	0.5
tgt	- 2	0.67	∞

increase, otherwise the problem becomes infeasible due to the current capacities on the arcs from a to $(1, 2, 3)$, all other things remaining equal.

Suppose $b_a = 2$, implying that the course of action can only affect two of the three initial target individuals. The new objective function value is 50, using as expected a different basis; only 1 unit of flow makes it to individual 10. If the course of action is only able to influence one of the three individuals (i.e., letting $b_a = 1$), both the objective function value and the basis change. The objective function value is 9 and neither one of the target individuals $\{9, 10\}$ are reached. This implies that another plan should be crafted, possibly seeking other individuals with more direct access to the target individuals, or ensuring that it is indeed possible to influence the initial set of individuals. The potential effects due to uncertainty in thresholds is worth considering. As indicated in Table 7.3, the threshold for individual 4, represented by b_4, may be increased (decreased) by no more than 1 (0.5) in order to maintain the current basis. Suppose the individual's threshold was actually more restrictive, thereby requiring more information, additional confirmation, or a signifi-

cant amount of persuasion by others before individual 4 would decide to promulgate the influence further into to network.

For example, letting $b_4 = -2$, the new objective function value is 64, with only one unit of flow reaching actor 10. Such a change suggests a more conservative response to influence from those reporting information to him. However, despite the change in this example, the course of action still provides enough influence to convince actor 4 to promulgate influence through the network and ultimately to one of the two target individuals. From here, it is up to the decision maker to decide if this potential result, influencing one of the two target nodes, is sufficient for the action's requirements and needs. If the value of this threshold were to change such that $b_4 = -3$, the problem still remains feasible, a change of basis occurs, and actor 4 prevents the propagation of influence to the target individuals. If the value of this threshold were such that $b_4 = 0$, the problem remains feasible, and both target nodes are fully influenced with $b_{tgt} = -2$.

All of the solutions presented thus far push at least some amount of influence through actor 4. An alternative path exists via arc $(1, 6)$ that can reach individual 10 and avoid actor 4. However, given the objective to minimize cost, this path tends to be avoided. This observation, as well as indications of uncertainty in the cost data, necessitates the investigation of the cost coefficient associated with this arc and any others that may be in question.

7.4.3 Changes in Risks

Recall that the cost coefficients $c_{(i,j)}$ may account for the perceived operation or personal risks of communication, actual costs of communication, and so forth. If the overall objective seeks to minimize cost, as in the example, these estimates play a major role in how much and along which paths influence flows through the social network. To assess the implications of changes in costs, an approach similar

Table 7.4: Cost Coefficient Analysis

$c_{(i,j)}$	Current	Allowable Increase	Allowable Decrease
$c_{(1,4)}$	8	0	3
$c_{(1,6)}$	20	3	0.33
$c_{(2,4)}$	6	∞	0
$c_{(3,4)}$	2	5	31.67
$c_{(4,5)}$	9	0.33	12.67
$c_{(4,6)}$	15	∞	3
$c_{(4,7)}$	24	∞	13.33
$c_{(5,6)}$	10	∞	29.83
$c_{(5,7)}$	6	0.25	9.50
$c_{(6,10)}$	6	∞	0.33
$c_{(6,11)}$	5	12.66	3.33
$c_{(7,8)}$	4	15	3
$c_{(7,9)}$	9	3	1.00
$c_{(7,10)}$	7	0.33	∞
$c_{(8,9)}$	8	∞	3
$c_{(9,tgt)}$	10	∞	1.00
$c_{(10,tgt)}$	9	3	∞
$c_{(11,10)}$	4	∞	3.33
$c_{(a,1)}$	6	0	∞
$c_{(a,2)}$	8	∞	0
$c_{(a,3)}$	7	5.00	∞

to varying the right-hand sides is taken. Given the current solution, the ranges of allowable change for cost coefficients are provided in Table 7.4.

Considering the cost estimate for $c_{(1,6)}$, the data in Table 7.4 suggests that a relatively small change (a reduction of ~ 0.33) in $c_{(1,6)}$ may result in a change in basis. Replacing this value with $c_{(1,6)} = 19$, the new solution (i.e., new basis) has an objective value of 92.66, with total flow to the *tgt* node improved to 1.64.

For some of the cost coefficients, such as $c_{(1,6)}$, a relatively small change may result in a new basis and therefore a new solution. This type of sensitivity suggests that multiple outcomes regarding flow within the network are possible, given even the slightest uncertainty in input data. Therefore, careful consideration in determining the course of action and its possible consequences should be made, particularly

with respect to the costs associated with interpersonal communication within the social network of interest. Since these costs are based upon the individuals' own perspectives, estimates of this particular parameter are potentially the most difficult to obtain via simple observation or surveillance. Consequently, courses of action that are robust or insensitive to changes in these coefficients would be preferred over those that are not. It may also suggest that an approach that lowers an individual's perceived cost may prove effective in directing the flow of influence.

The last remaining model input considered for post-optimality analysis is that of the arc capacities, which serve as estimates of social closeness. The closer two individuals are the stronger the relationship, resulting in a potentially greater degree of influence exchange between them. These values are also subject to uncertainty.

7.4.4 Changes in Social Closeness

The problem formulation that was optimized initially took advantage of a feature that accounts for the upper bounds associated with the arc capacities. This is given as simple upper bound (SUB) for any variable $x_{(i,j)}$. This facilitates performance (time to solution) by implementing the generalized upper bounding technique, but also eliminates the ability to use post-optimality information often provided by the software [Ahuja et al., 1993, pg. 666-7]. Specifically of interest is the evaluation of changes in the upper (or lower if applicable) bounds for the arc capacities. An easy way to remedy this is through the inclusion of this capacity constraint within the constraint set, as opposed to using the SUB function. Adding the constraint $x_{(i,j)} \leq u_{(i,j)}$ provides an opportunity to then investigate the implications of uncertain data regarding the estimates of social closeness used within the social network in a manner identical to that of changes in the thresholds, sources, and sinks, without changing the solution results. Accounting for all current arc capacities in the constraint set of the original GNF formulation for the network in Figure 7.7, the post-optimality results for all arcs with non-zero flow are shown in Table 7.5.

Table 7.5: Arc Capacity Analysis

Variable	Flow	Current Capacity	Allowable Increase	Allowable Decrease
$x_{(1,4)}$	1	2	∞	1
$x_{(2,4)}$	1	3	∞	2
$x_{(3,4)}$	1	1	0	0
$x_{(4,5)}$	2	7	∞	5
$x_{(5,7)}$	2.66	5	∞	2.33
$x_{(7,9)}$	0.33	6	∞	5.67
$x_{(7,10)}$	1	6	∞	5
$x_{(9,tgt)}$	0.33	1	∞	0.67
$x_{(10,tgt)}$	1	1	0.33	0.67
$x_{(a,1)}$	1	1	∞	0
$x_{(a,2)}$	1	1	1	0
$x_{(a,3)}$	1	1	∞	0

The allowable ranges indicate the change in the arc capacity that can be tolerated and still maintain the current basis. This type of analysis offers a means to assess whether or not the current estimate of social closeness between two individuals plays an important role in the current solution. The smaller the allowable range, the more important an accurate assessment of social closeness is required. Note that the variables or arcs that comprise the social network, as opposed to those emanating from node a or going to node tgt, are the main concern in this setting. Arcs exhibiting relatively small ranges, and therefore sensitivity, should be of particular interest, the values of which should be verified by additional intelligence information as necessary.

Once expressed in a network flow context, influence flowing through a social network lends itself to an array of post-optimality analyses. Given the likely imprecise nature of the inputs, particularly in a military or political setting, the ability to conduct post-optimality analysis is critical in attempting to model influence and behavior. Knowing the range of applicability of a solution or a parameter provides the decision maker with an estimate of the robustness of a course of action.

7.5 Summary

Social science literature has developed numerous theories and measures in order to better understand human interaction and its consequences. Numerous explicit and implicit connections have been suggested between social and physical networks; using the analogy of influence as a pseudo-physical commodity, these connections facilitate the study social networks via the generalized network flow problem.

While improvements are desirable in order to improve the quantification of social phenomena serving as inputs to this methodology, the operations research tools, in this case the simplex method, provides an advantageous byproduct of a variety of post-optimality analyses. As several aspects of the notional network appeared to be sensitive to changes, the example reiterates the need for accurate, objective estimates of network dynamics. It is posited that all of these capabilities will culminate into a methodology to evaluate and develop courses of action for influencing social networks.

Of course, the ultimate decision will remain with the decision maker. This approach provides the information campaign planner a means to investigate alternative courses of action, and perhaps to aid in developing intelligence requirements where the sensitivity and parametric analysis suggests. Given this base, a number of other approaches and variations can be investigated and modeled.

VIII. Case Study

8.1 Chapter Overview

The purpose of this chapter is to illustrate, compare, and contrast analysis techniques developed during the course of this research. Consequently, this chapter demonstrates some of the steps involved in the processes when applying these concepts, as well as highlighting the need to carefully consider the associated underlying assumptions. The data used, drawn from a dated open-source study, are merely for illustrative purposes. The data are subjected to RBAP, KPP-2, network flow centrality, and generalized network flow centrality analysis. The analysis is notional in nature and is intended to be illustrative rather than being interpreted as an actual operational study.

8.2 Data Description

The data analyzed within this chapter comprise a network of 48 individuals with known or alleged ties to the Jemaah Islamiya terrorist network, commonly referred to as JI. These members are a subset of the open source Al Qaeda network data developed and analyzed by Sageman [2004], and were selected by subject matter experts due to their affiliation with JI [Clark, 2005, pg. 5-1]. It is assumed that link information within this data set are associated with symmetric ties between two given individuals. With the exception of RBAP analysis, the assumption of symmetry could easily be relaxed. While the existence of negative ties are certainly of interest due to their potentially detrimental effects upon the strength of personal ties, as well as their susceptibility to exploitation, the mathematical characterization of this case study assumes existing ties are always positive in nature. The member names and corresponding identification numbers are provided in Appendix O.

JI terrorist cells predominantly span Southeast Asia, generally operating within Indonesia, Malaysia, and the Philippines [U. S. Department of State, 2005, pg. 33, 101]. Membership estimates vary between the hundreds and the thousands; the organization has been confirmed to be "responsible for numerous high-profile bombings" such as the hotel bombings in Bali (2002) and Jakarta (2003) [U. S. Department of State, 2005, pg. 101]. Al Qaeda and JI are directly linked through Riduan bin Isomoddin (also known as Hambali), who was the leader of JI and the Southeast Asia operations chief of Al Qaeda until his capture in 2003 [U. S. Department of State, 2005, pg. 101].

If the different contexts contributing to the relationships of the 48 case study members were ignored and simply characterized as "a tie exists or not," the resulting social network is shown in Figure 8.1. Note, however, that these ties have evolved from one or more contexts or situations. Specifically, these relationships have resulted from at least one of six relations discernable by (open source) intelligence information. Although there are a variety of relations that could be discerned, such as those offered by Hite, or possibly geophysical location networks, the relations extracted by Sageman include *discipleship*, *worship*, *familial*, *relative*, *friend*, and *acquaintance* networks. For the purposes of this example, it is assumed that all members under study remain at large and actively involved within the terrorist organization. While this is not the actual case, it does not reduce the illustrative nature of the case study. Clearly there are temporal aspects that must eventually be addressed, as they may offer some insight into how the network may be evolving, growing, shrinking, and so forth. The approach described here does not preclude including temporal effects as changing snapshots in time. The various contexts comprising these relationships are depicted in Figures 8.2 through 8.7 and were developed using the visualization tool within the Organizational Risk Analyzer (ORA) [Carley, 2006]. Nodes displayed on the left-hand side of each graph have no known relationships among other individuals within that context. Technically, they do not qualify as isolates within that

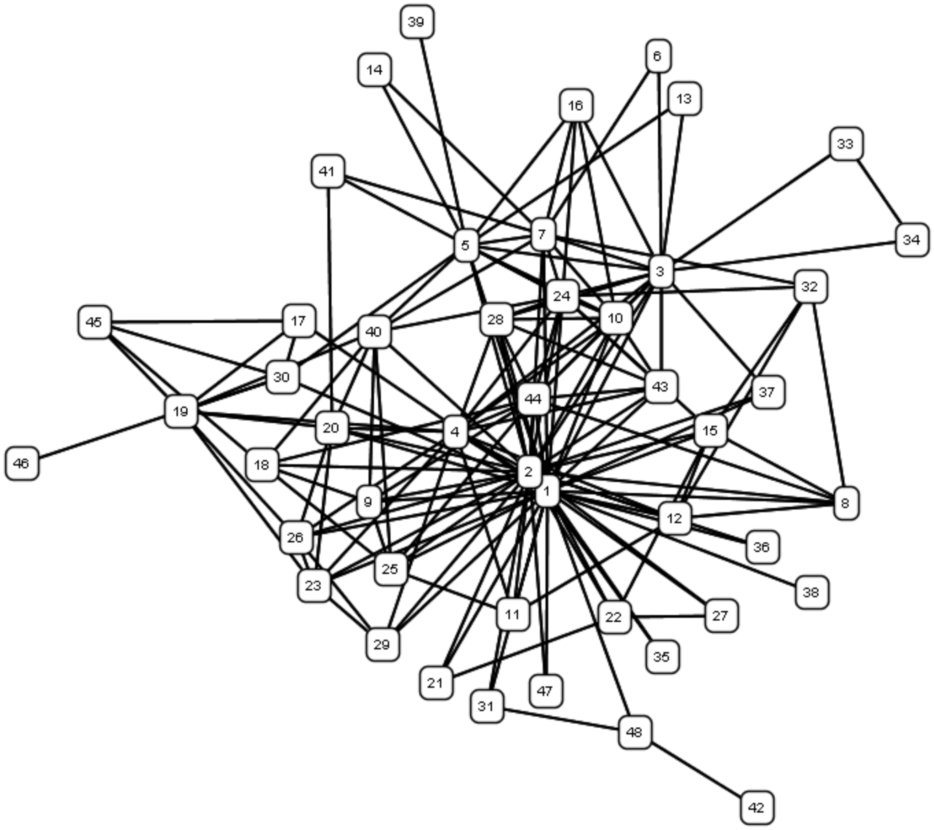


Figure 8.1: JI Combined Network for 48 Core Members

context's graph. Their appearance is only meant to highlight the separated individuals and maintain the intent of the layered concept that considers all actors within the organization of interest.

The network shown in Figure 8.2 illustrates the affiliation of *discipleship*, which clearly shows the prominence of Baasyir (1) and Sungkar (2). Due to limited descriptive information associated with the data, it is unclear as to the significance of cluster of actors Zulkarnaen (24), Dulmatin (32), Yunos (16), Syawal (7), Hambali (3), Iqbal (5), and Sufaat (13). However, Hambali (3) and Zulkarnaen (24) have served high-level leadership roles, leader and military chief, respectively [Abuza, 2006, pg. 4]. In addition, all of the individuals within this cluster have known friendship ties, and some share familial ones. Bassyir (1) and Sungkar (2) are the founders of the Is-

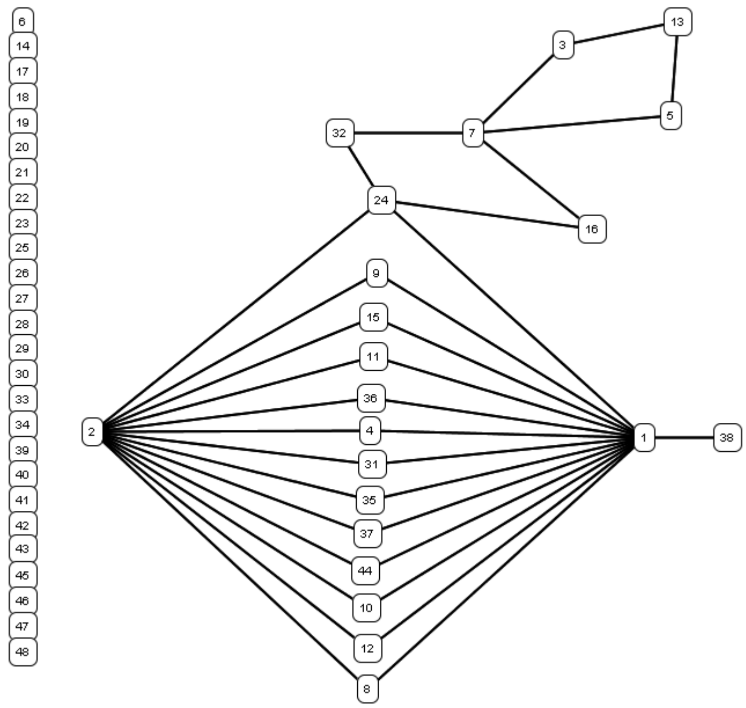


Figure 8.2: JI Discipleship Network

lamic boarding schools that ultimately resulted in the initial JI membership analyzed within this case study [Sageman, 2004, pg. 113].

The *worship* network in Figure 8.3 illustrates some overlap between the two primary teachers, Baasyir (1) and Sungkar (2), as well as indirect contacts. If possible, a multigraph depicting which individuals attended the mosques purported to recruit members for the Jihad would be valuable information to supplement the overall network structure. The social practice of worship, sharing, or learning extremist interpretations of Islam plays a significant role in the "the process of affiliation to the Jihad" [Sageman, 2004, pg. 114].

Figures 8.4 and 8.5 shows the kinship relations among the selected individuals, where the relation is either through marriage (the *relative* network) or familial (the *family* network), respectively [Sageman, 2004, pg. 112]. The last two networks in Figures 8.6 and 8.7 essentially capture two different types of social interaction. The

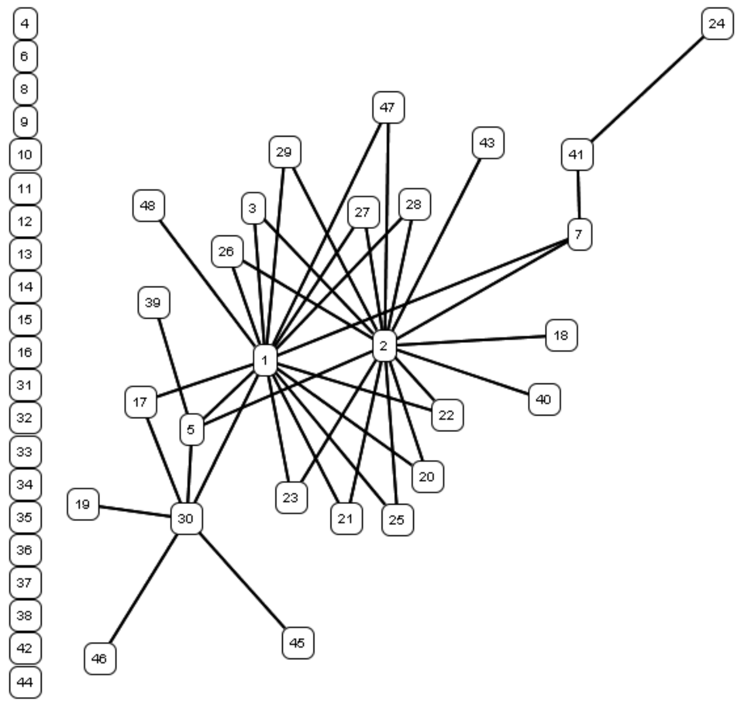


Figure 8.3: JI Worship Network

first, *friendship*, suggests a stronger type of social bond than the latter, *acquaintance*. These could correspond to the strong and weak ties discussed by Granovetter [1973], respectively.

Note that each of these contexts could be perceived as the layers defining the interpersonal relationships of these individuals, contributing, in varying amounts, to the strength of ties among them. Exactly how these layers are combined to ascertain the strength values are likely dependent upon the organization of interest. Both the similarity and decision theoretic approaches are illustrated in Section 8.5.

In addition, data capturing individual characteristics were also provided in the sample data; the categories of data are shown in Table 8.1. This information serves as inputs to the gain multiplier methodology described in Chapter VII, which is demonstrated in Section 8.5.2.

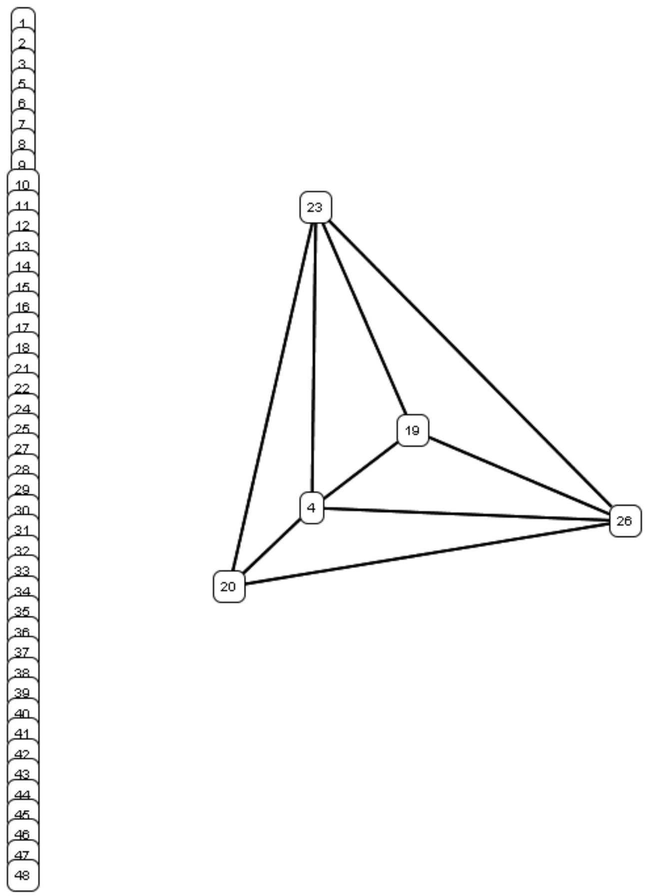


Figure 8.4: JI Relative Network

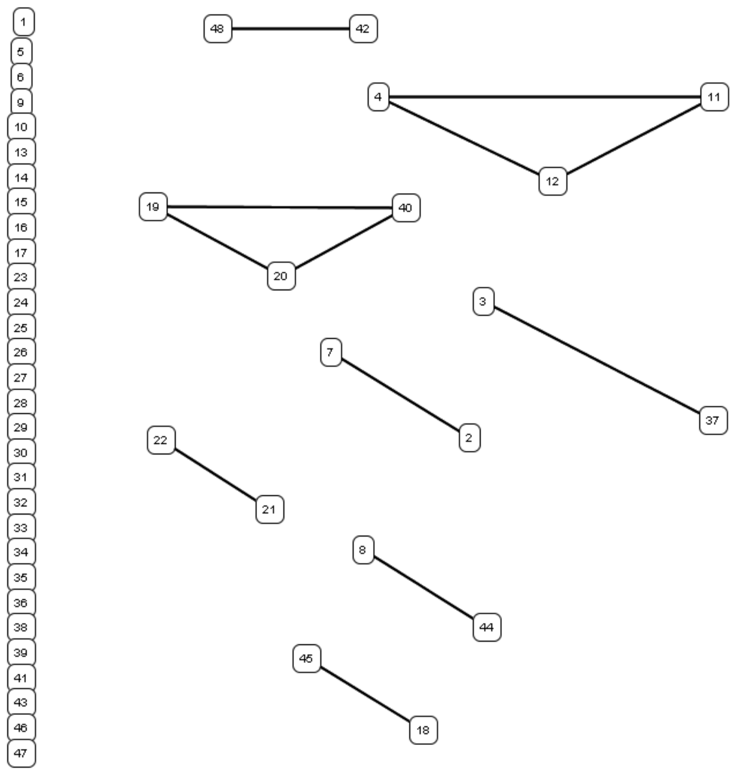


Figure 8.5: JI Familial Network

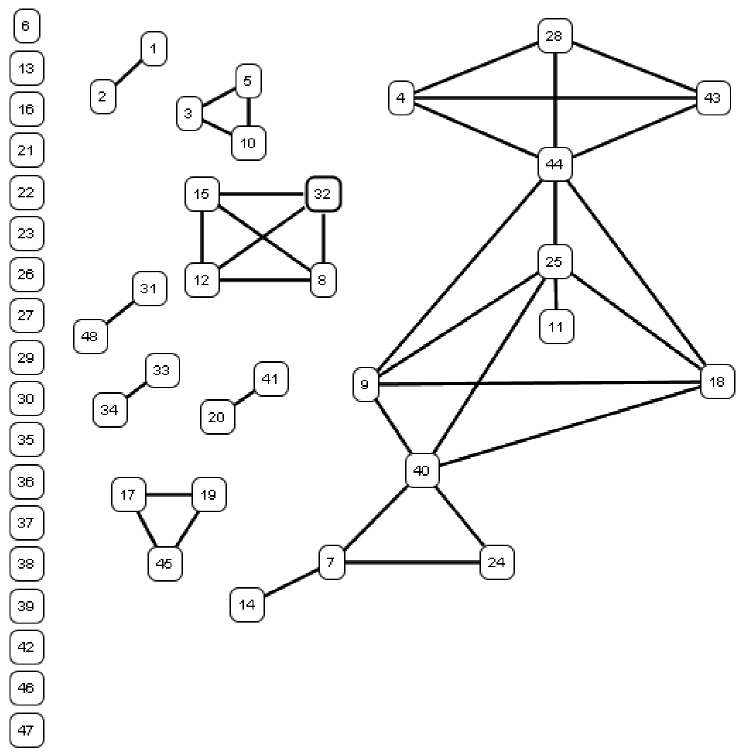


Figure 8.6: JI Friendship Network

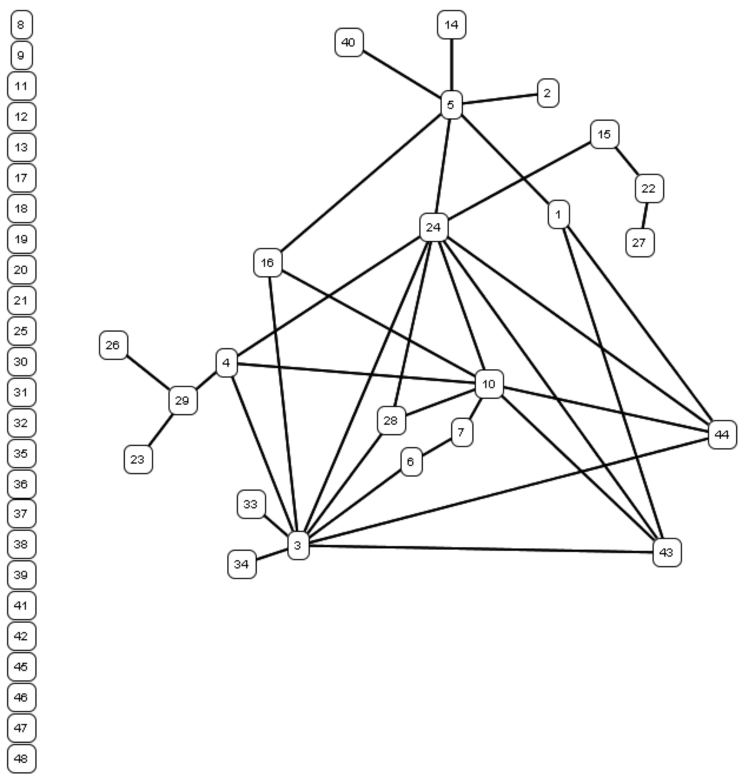


Figure 8.7: JI Acquaintance Network

Table 8.1: Actor-Specific Data [Sageman, 2004]

Individual Attribute
Short Name
Full Name
Date of Birth
Country of Birth
Clump (Regional or organizational grouping)
Youth National Status
Family Socio-economic Status
Religious background
School Attended
Educational Achievement
Type of Education
Occupation
Marital Status
Children
Social Background (Criminal)
Role (Position) in Organization
Year joined the Jihad
Age joining the Jihad
Place joined the Jihad
Country Joined the Jihad
Fate
Year left

8.3 RBAP Analysis

Considering that very limited data is available at the initial phases of analysis and network discovery, the RBAP centrality measure is first applied in order to demonstrate the screening process. Recall that the RBAP measure offers a means to do a preliminary screening of individuals, determining which individuals have an advantageous position based upon network topology uncovered to this point. Since the relations of interest are assumed to be symmetric, the application of RBAP yields a centrality measure of actor position. Note that α represents the attenuation of influence or information as a function of path distance. Therefore, RBAP with $\alpha = 0$ is equivalent to degree centrality, offering a local perspective on centrality. RBAP with $\alpha = 1$ measures the number of other actors that can be effectively reached, offering a global perspective of centrality.

Both sets of results are of interest, but for different reasons. The locally central individuals are able to potentially directly influence the greatest number of other individuals, whereas the globally central individuals may serve as either advisors to the locally central members or perhaps as liaisons to other organizations or regions of operation. The relation between RBAP scores and actor position in these relatively small networks may appear obvious. However, when dealing with new data, characterizing networks of hundreds to thousands of individuals, a screening tool of this nature would facilitate further analytic efforts.

Using the RBAP sensitivity analysis procedure provided in Appendix B, comparisons can be made as α is varied between 0 and 1 for a given network. Note that, due to the path-based nature of RBAP, it is recommended that the measure be applied only to connected networks such as the *combined*, *discipleship*, *worship*, *relative*, and *acquaintance* networks.

Comparisons between the top five high-scoring individuals at both $\alpha = 0$ and $\alpha = 1$ were made for the *worship*, *discipleship*, and *combined* networks. The former

two were selected due to the significant role that *worship* and *discipleship* activities play in indoctrinating individuals and building trust among members [Sageman, 2004, pg. 114]. The latter network was selected because a conglomerate type of network is likely, assuming that the type of data immediately available for analysis is derived from an automated intelligence process that must assess relationships from a distance.

Applying RBAP to the *worship* network (Figure 8.3), the results for $\alpha = 0$ and $\alpha = 1$ are shown in Table 8.2. With this in mind, Baasyir (1), Sungkar (2), and Maidin (30) have the highest local RBAP scores due to their central and well-connected position within the *worship* context. This result for the first two actors is not surprising, given this context; no descriptive information is available on Maidin (30) in this data set. These results suggest that increased investigative efforts focusing on Maidin (30) would be warranted.

With $\alpha = 1$, the measure ranks Zulkarnaen (24), Thomas (48), and Faiz (18) highest, respectively, all of whom joined the organization after Baasyir (1), Sungkar (2), and Maidin (30). Zulkarnaen (24) has the lowest closeness centrality among all others, due to his peripheral location; however, his relatively close access, as measured by distance in links, to the most central members results in the highest global RBAP score when attenuation of information or influence as a function of path distance in links is assumed to be non-existent (i.e., $\alpha = 1$). This suggests that, despite his peripheral location, Zulkarnaen (24) may be an influential member of the group, a potential connection to another, entirely separate organization, or perhaps a promising access point to the network. These conditions are indeed the case, as Zulkarnaen (24), purportedly the chief of military operations, has assumed the overall leadership responsibilities of Baasyir (1).

Turning to the *discipleship* network (Figure 8.2), it is evident that Baasyir (1) is a central actor, likely due to his role in originating the group. Interestingly, Sufaat (13) scores the highest with $\alpha = 1$. Sufaat (13), joining about 9 years after the group

Table 8.2: RBAP (Worship Network)

$\alpha = 0$		$\alpha = 1$	
Actor	Score	Actor	Score
Baasyir (1)	16	Zulkarnaen (24)	19.14
Sungkar (2)	16	Thomas (48)	17.64
Maidin (30)	6	Faiz (18)	17.58
Iqbal (5)	4	Hafidh (40)	17.58
Syawal (7)	3	Rusdan (43)	17.58

Table 8.3: RBAP (Discipleship Network)

$\alpha = 0$		$\alpha = 1$	
Actor	Score	Actor	Score
Baasyir (1)	14	Sufaat (13)	20.5
Sungkar (2)	13	Marzuki (38)	15.36
Syawal (7)	4	Baasyir (1)	15.26
Zulkarnaen (24)	4	Mukhlas (4)	14.96
Hambali (3)	2	Ghozi (8)	14.96

was purportedly formed, became increasingly religious throughout his tenure in JI, studying with senior members and "was the host for the Kuala Lumpur al Qaeda conference leading to the USS Cole bombing and the 9/11 operations" [Sageman, 2004, pg. 112].

Applying RBAP to the *combined* network, formed by taking a Boolean sum across all network layers, the top 5 ranked individual scores are shown in Table 8.4. As one may expect, Baasyir (1) and Sungkar (2) are the most central from a local (degree) perspective. Marzuki (38) is ranked highest due to his peripheral location and direct connection to Baasyir (1). Marzuki (38) is purportedly the chief financier for JI and remains at large [Meng, 2004].

Using the *combined* network shown in Figure 8.1, the correlations between RBAP, at $\alpha = 0$ and $\alpha = 1$, and all other standard centrality measures are provided in Table 8.5. As expected, degree centrality and RBAP at $\alpha = 0$ are perfectly correlated. RBAP at $\alpha = 1$, as would also be expected, does not correlate well with

Table 8.4: RBAP (Combined Network)

$\alpha = 0$		$\alpha = 1$	
Actor	Score	Actor	Score
Baasyir (1)	32	Marzuki (38)	33.39
Sungkar (2)	30	Baasyir (1)	33.34
Hambali (3)	16	M. Yunos (35)	31.94
Mukhlas (4)	15	Naharudin (36)	31.94
Zulkarnaen (24)	15	Roche (47)	31.94

Table 8.5: RBAP Correlations to Other Measures

	Degree	Closeness	Betweenness	Eigenvector	RBAP ($\alpha = 0$)
Closeness	0.92	–	–	–	–
Betweenness	0.87	0.77	–	–	–
Eigenvector	0.95	0.95	0.72	–	–
RBAP ($\alpha = 0$)	1	0.92	0.87	0.95	–
RBAP ($\alpha = 1$)	0.28	0.4	0.4	0.32	0.28

the other measures, since it is capturing a process significantly different from those assumed in traditional measures.

Since none of the network layers share exactly the same set of actors, only qualitative assessments may be made across network layers. For example, Baasyir (1) is in the top 5 most central actors from both a local and global perspective in all three of the network layers analyzed, with the exception of the *worship* network at $\alpha = 1$, where he scores sixth. Although he is now known to be one of the co-founders of JI, the screening approach offered by RBAP, used early in the investigative process, clearly signals him as a potential actor of interest.

8.4 Key Player Analysis

All of the network layers within this case study are relatively small, which often results in relatively small domatic numbers. For example, examining the *combined* network from Figure 8.1, the domatic number given a maximum reach of 2 ($\delta_2 = 1$).

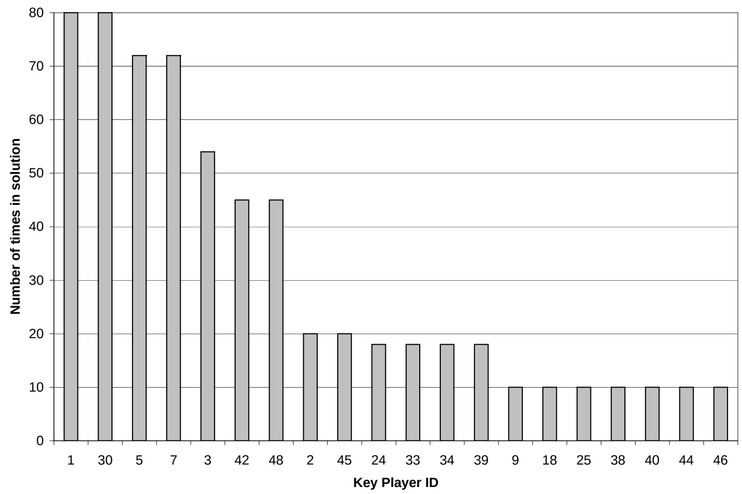


Figure 8.8: NR2 (Combined Network)

The only solution to this problem is Baasyir (1). Consequently, if a maximum reach of 2 is permitted, Baasyir is the key player from the NR2 problem perspective.

If key players were limited to influencing only those actors adjacent to them (NR1), the domatic number increases to $\delta_1 = 7$. All 80 optimal solutions were generated. Recall that Borgatti's heuristic generates only a single solution. Developing all the optimal solutions provides a wider set of targeting options. Figure 8.8 provides the histogram of the number of times a particular actor comprised an NR1 solution. Actors Baasyir (1) and Maidin (30) appear in every optimal solution. Iqbal (5) and Syawal (7) appear in 72 of the 80 optimal solutions. Hambali (3), one of the leaders of JI, appeared in 54 of the 80 optimal solutions [Sageman, 2004, pg. 44, 138].

Another effect due to small size and connectivity of this network is that the key players are, in this case, primarily comprised of the leadership. Considering the general concept of the KPP-2 concept, to influence a network in an efficient manner, this coincides with such a theoretical goal. The practical goal, however, must consider the fact that convincing the adversarial leadership to promulgate influence to our own Nation's benefit is unlikely to occur. Therefore, it may be of interest to designate individuals not eligible to serve as a key player. As discussed in the key player

methodologies developed in Chapter V, this is easily accomplished via additional constraints in the mathematical program. For example, suppose that the decision maker still wanted to influence the entire network, relying upon a reach of no more than two steps away from a key player. In addition, the solutions must require that Baasyir (1), Maidin (30), Iqbal (5), and Syawal (7) are not selected. Given that $x_i = 1$ if actor i is selected as a key player, 0 otherwise, the addition of the constraint $[x_1 + x_{30} + x_5 + x_7 = 0]$ achieves the desired effect. With this constraint in place, the domatic number increases to $\delta_2 = 9$, and there are only six optimal solutions. The histogram of key player occurrences within the solution is provided in Figure 8.9. Hence, tradeoffs exist that can be explored via the proposed methodology.

Similar constraints could be incorporated within the other variations of KPP-2 shown in Table 5.6. Note that seven of the nine individuals required to satisfy the NR2 problem in this setting appear in all six solutions; these include Sungkar (2), Hambali (3), Dwikarna (14), Azahari (24), Marzuki (38), Kastari (39), and Khalim (46). The next step would be to assess the likelihood of successfully co-opting all of these actors simultaneously. If this is not possible, then the fractional key player problems could be used to assess further tradeoffs between access to potential players and the effectiveness of the planned information operation.

If the key players were required to influence or contact all others within all contexts, the key player algorithms cannot (and should not) simply be performed on an aggregate network, as any given player may not be connected to the same actors in all layers. An example of such a requirement could include the assignment of key players to convey influence within the given contexts that form their personal relationships. If this requirement is levied upon all layers, a reinforcing effect results, ensuring that each actor is influenced by a key player in all contexts. The domatic number for this multi-graph is 16 actors. Figure 8.10 depicts the histogram of player occurrence within the 20 optimal solutions. There are 10 players that occur in every one of the optimal solutions. Again, this indicates a possible need to evaluate the

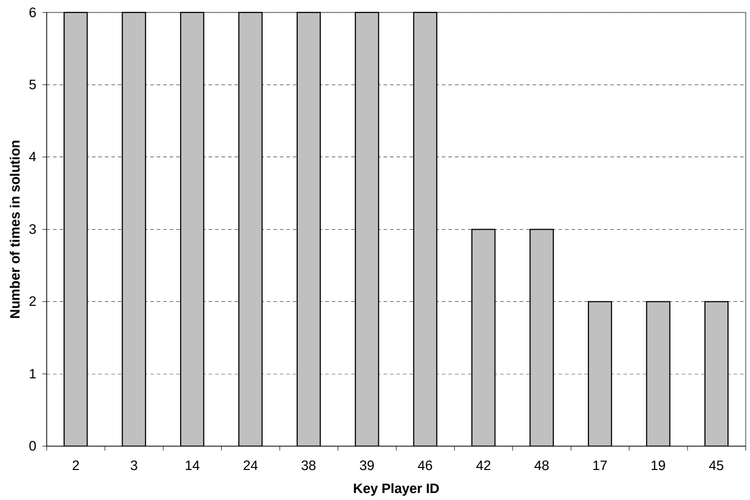


Figure 8.9: NR2 without Leaders (Combined Network)

operational tradeoffs if all of these actors cannot be convinced to serve as a key player simultaneously. Note that the constraint matrix must be generated in a slightly different manner due to varying isolates among the layers. Appendix M describes the steps required to perform this type of analysis and provides accompanying MATLAB code.

8.5 Network Flow Analysis

To demonstrate some of the main methodologies developed within this research, application of the network flow centrality measure by Freeman et al. [1991] to variants of the aggregated network are compared. Three cases are explored: (1) intelligence information is strictly limited to knowledge of existing ties; (2) intelligence information has knowledge of the existence and composition (layers) of interpersonal ties; and, (3) intelligence information consists of the composition of known ties, individual attributes, and subject matter expert opinion regarding the relative influence each individual may exert over others.

To investigate the first case, the network flow centrality procedure outlined in Appendix K was applied to the sociomatrix corresponding to the *combined* network

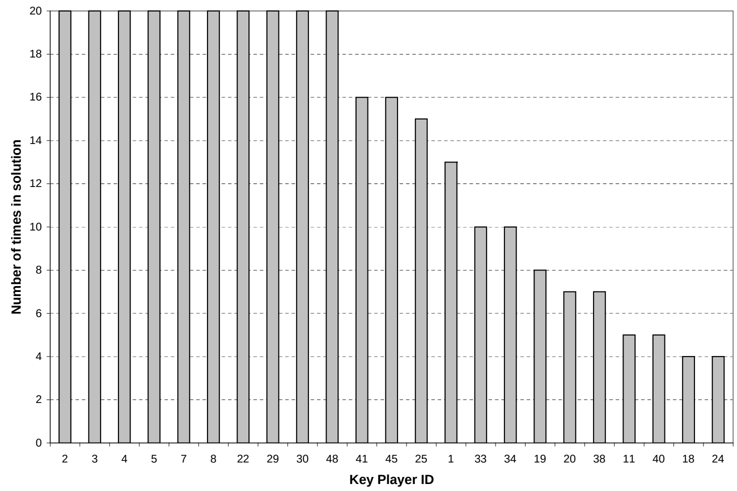


Figure 8.10: NR1 for all layers simultaneously

in Figure 8.1. The results are shown in column 1 of Table 8.6. To investigate the remaining cases, several assumptions and data must be considered.

8.5.1 Arcs

In order to take advantage of all possible information, estimates of interpersonal tie strength serve as arc capacities, similar to the approaches described by Renfro [2001] and Clark [2005]. Assuming that this information is derived from the contexts comprising the known relationships, recall the two methods of measurement developed in Chapter VI: the similarity-based and decision-theoretic approaches.

The similarity approach begins with combining the node-edge adjacency matrices for each layer into one, conceptually shown in Equation 8.1, which then serves as input to the Jaccardian similarity measure provided in Appendix I.

$$NE_{all} = [NE_{family}\,|NE_{friend}\,|\ldots|\,NE_{discipleship}] \tag{8.1}$$

The interpersonal tie strength between actors i and j derived from this Jaccardian similarity approach is denoted $s_{JS}(i, j)$. The strongest tie strength observed given

this data set is 0.33. The network flow centrality results are provided in column 2 of Table 8.6.

The decision-theoretic approach as applied to the case study data is based upon the model shown in Figure 6.6. For the purposes of this example, it is assumed that the *discipleship* and *worship* relations can be classified as significant events contributing to tie strength, as these are often contributing factors to the recruitment and assimilation into JI [Sageman, 2004, pg. 114]. Consequently, the *familial*, *relative*, *friendship*, and *acquaintance* networks are classified as relationships contributing to the intimacy aspect of tie strength. For the time component, data capturing when each pair-wise relationship was confirmed was not available; a proxy was developed. The year each member joined the organization was available for most (44 of the 48) members. For missing data, the average value for year joined (1991) was arbitrarily selected. Let $t_{lastobservation}$ denote the year of the last observation or tie discovery (2001). Let $t_{join}(i)$ denote the time that a given actor i joined the organization. Given two individuals i and j, if a tie between them was known to exist in any one of the layers, the estimated amount of time the tie was assumed to correspond to

$$t(i,j) = t_{lastobservation} - max[t_{join}(i), t_{join}(j)] + 1. \tag{8.2}$$

This approach serves as a proxy to estimate the possible length of time a tie existed, based upon limited information. The scores range from 0 to 13. In order to derive a value for the decision model, a linear function was assumed, and each score was simply normalized by dividing by the time of the longest established tie (13 years). Any appropriate form may be used, however.

Again, this ignores the fact that some individuals left the organization for one reason or another throughout the period between 1989 and 2001. In addition, although the original measure capturing the time data pertinent to a specific tie is preferred, this example is for illustrative purposes only. Finally, for illustrative purposes

the model weights were evenly distributed; for example, let $w_{time} = w_{SE} = w_I = 1/3$, $w_d = w_w = 1/2$, and so forth. Then, using the conceptual model presented in Equation 6.9, the decision-theoretic measure of tie strength, denoted s_{DA}, is derived by

$$s_{DA}(i,j) = w_{time} v\left[t(i,j)\right] + w_{SE}\left[w_d X_d + w_w X_w\right] + w_I\left[w_{fr} X_{fr} + w_{fa} X_{fa} + w_r X_r\right]. \tag{8.3}$$

Standard weight elicitation approaches would be used in a specific analysis, tempered by available time and experts. Using the evenly distributed weights, available network data, and the proxy for the amount of time spent within the relationships, the strongest tie strength observed given this data set, assumptions, and approach is 0.583.

From Table 8.6, Baasyir (1), Sungkar (2), and Hambali (3) all share the top three positions in all variants of flow centrality: using only the adjacency matrix $\mathbf{X}$ as input to the network capacity, topologically accounting for similarity s_{JS}, and topologically accounting for how the different contexts theoretically contribute to the strength of interpersonal ties (s_{DA}). Individuals Iqbal (5), Syawal (7), Zulkarnaen(24) and Mukhlas (4) all vie for the next highest centrality scores. Given that similar results were found for these men using the RBAP and KPP-2 techniques, this suggests that placing increased investigative efforts upon them could be beneficial. The next section repeats the network flow centrality, but also accounts for sender-receiver specific effects upon the amount of information or influence traveling among individuals.

8.5.2 Gains and Losses

To estimate the gain or loss of influence or information due to the sender-receiver effects, the actor-specific data elements shown in Table 8.1 were reviewed for their relevance to a logistic regression model as discussed in Section 7.2.3. All but the names, fate, and year the individual left the organization were considered

Table 8.6: Network Flow Centrality (Top 10 Individuals, Without Gains)

X		Jaccardian		Decision Analysis	
ID	Centrality	ID	Centrality	ID	Centrality
1	0.176	1	0.112	1	0.176
2	0.136	3	0.095	2	0.146
3	0.096	2	0.091	3	0.083
5	0.056	19	0.073	7	0.057
7	0.053	5	0.06	5	0.056
24	0.041	7	0.058	24	0.047
4	0.041	30	0.056	4	0.042
19	0.035	22	0.05	30	0.037
30	0.031	15	0.043	19	0.036
44	0.029	45	0.04	44	0.031

Table 8.7: Categories of Leadership [Clark, 2005, pg. 5-4]

Leadership Level	Description of Members
1	Emir Types (Senior Leaders/Founders)
2	Trusted Second Tier/ Key Counselors and Facilitators / Leadership Council
3	Regional/District Leaders / Key Operatives / Unit Commanders / Liaisons
4	Operatives who provide support or followers who often risk arrest, physical injury or death (i.e., execute missions) / foot soldiers

as potential candidates for inclusion as predictor variables. The binary response variable indicated the charisma of an individual.

For the purposes of this research, charisma represents the influential nature of an individual based upon either their authoritative roles, their personal characteristics, or both. The response was based upon an initial classification of individuals by subject matter experts, shown in Table 8.7. To provide a demonstration of the methodologies, it was assumed that if an individual i was in classified as 1 or 2, then $y_i = 1$, and 0 otherwise.

The initial list of predictor variables was pared down for a variety of reasons. For example, the *Age Joined* and *Date of Birth* were highly negatively correlated

(-0.716); as a result, only one of these, *Age Joined*, was selected as a possible predictor variable. The *Clump* characteristic was the same value for all actors within the sample set, thereby providing little or no explanatory power. Ultimately, the predictor variables *Year Joined* and *Age Joined* were selected due to their ability to predict the response variable while dealing with missing data.

Operationally, each of these personal attributes could represent the authoritative and personality aspects of source-dependent influence. For example, the longer an individual has been a member, the more likely he has participated in operations, built trust among members, and risen to leadership positions. The age at which an individual joins, and the effect of this attribute upon that individual's ability to persuade others is likely culturally specific. For this case study, it is assumed that elders, particularly religious ones, are well respected and revered due to their perceived wisdom and experience. Consequently, if two members join at the same time, the older one will be more influential and vice versa. Alternatively, if two members had joined JI at the same age in their lifetime, the one who has more time in the organization will be more influential, and vice versa. The process used to develop pair-wise gains is documented in Appendix N.

Let X_0 denote the intercept, X_1 denote the year a given individual joined JI, and X_2 denote the age when membership began. Using the results in Table 8.8, the estimated probability that a given individual i with characteristics X_1 and X_2 is charismatic (i.e., $y_i = 1$) is

$$\hat{\pi}_i = \frac{e^{(1087.8 - 0.548X_1 + 0.099X_2)}}{1 + e^{(1087.8 - 0.548X_1 + 0.099X_2)}}. \tag{8.4}$$

Once the estimated probabilities are calculated for each actor, Equation 7.4 is used to determine the gain multipliers g_{ij} for all actor pairs with a link in the contexts or layers of interest. This information can then be used as input to the generalized network flow centrality measure developed in Chapter VII. Table 8.9

Table 8.8: Logistic Regression Results

Variable	Coefficient	Standard Error	χ^2	p-value
Intercept	1087.8	626.7	3.013	0.083
Year Joined	-0.548	0.313	3.033	0.082
Age Joined	0.099	0.059	2.733	0.098

Table 8.9: Network Flow Centrality (Top 10 Individuals, With Gains)

X		Jaccardian		Decision Analysis	
ID	Centrality	ID	Centrality	ID	Centrality
1	0.17	3	0.14	1	0.16
3	0.13	1	0.11	2	0.12
2	0.12	30	0.09	3	0.11
5	0.09	2	0.08	5	0.08
30	0.07	5	0.08	30	0.08
4	0.06	15	0.07	4	0.06
7	0.05	19	0.07	7	0.05
15	0.05	7	0.06	15	0.05
24	0.05	22	0.06	24	0.05
19	0.04	4	0.05	19	0.04

compares the top 10 highest ranked individuals with respect to generalized network flow centrality. Baasyir (1), Sungkar (2), and Hambali (3) continue to rank highly due to their network location. However, the inclusion of personal attributes and their effect upon interpersonal influence brought Maidin (30) into the top three when the Jaccardian approach to tie strength is used. Interestingly, Maidin (30) was the leader of the JI members within Singapore, suggesting that looking at these networks from various perspectives may provide insight into organizational strengths, vulnerabilities, and individuals that should be subjected to either further intelligence scrutiny or military action [Ministry of Home Affairs, Singapore, 2002]

This new generalized network flow approach to centrality not only accounts for network topology, but also includes the effects upon information or influence due to the interaction of different individuals, no longer viewing the nodes as homogenous entities. Therefore, this is another option towards better models of social networks.

8.6 Summary

The measures related to arc capacity and gain factors are amenable for use in influence course of action analysis as presented in detail within Section 7.3. Overall, these techniques serve as flexible tools to process and evaluate new information (via RBAP), account for detailed information regarding the interpersonal relationships (via the information garnered from layers and individual characteristics), and explore potential influence courses of action and their efficacy. All of these methods are designed to serve the intelligence community in the process of knowing the adversary, and eliminating them.

IX. Conclusions and Recommendations

9.1 Overview

Recalling the Colonel Mathieu quote, "To know them means to eliminate them," this research sought to develop new theory and means of investigating non-cooperative networks, facilitating increased insight into network arrangement and evolution, with the goal of identifying potential opportunities for influence operations. Such opportunities lie within the ability to disrupt the effective operation and growth of these networks, or destroy them entirely. Although these adversaries can be affected in a number of ways, this research focuses upon either removing or mitigating an organization's most influential individuals, or finding susceptible points of entry and conveying information or influence that contributes to winning this war of ideas. This chapter provides a summary of the methodological and practical contributions of this dissertation, as well as recommendations for areas of future research.

9.2 Dissertation Contributions

This research examined non-cooperative networks from a number of perspectives, with data requirements ranging from single-layered (or simply aggregated) topological data to complex, multi-layered network data also requiring information characterizing the individuals themselves.

The study begins with the assumption that limited information that captured the dyadic interactions between individuals was available. The methodology offers a screening tool that lends itself to the data-sparse environment initially confronted by analysts. The reach-based assessment of position centrality measure is founded upon an extension of previously existing graph theoretic and social network analysis methods. To model the surreptitious communications among clandestine networks,

this new measure focuses upon the flow of information or influence along geodesics. Implementation of the attenuation concept, shared among several other centrality measures, is improved upon, yielding a capability to analyze the full range of possible options. This measure incorporates these theoretical aspects such that the technique is amenable to non-cooperative networks, is computationally attractive, and is freed from the analytic constraints inherent to similar social network analysis measures.

The next area of research improved upon the key-player problem (KPP-2) described by Borgatti [2003a]. This concept is generally viewed as a group centrality problem, which attempts to identify the most central group of people within a network rather than its most central actor. Applying mathematical programming techniques to this relatively new sociological problem yielded not only models equivalent to Borgatti's original problems and concomitant heuristic procedures, but techniques to investigate new aspects of the key-player problem as well. These extensions accommodate specific analysis requirements as well as other methodological constructs presented in other areas of this research. Examples of these extensions include the use of interpersonal tie strength values as input into the p-median variant of KPP-2. As evidenced by the variety of applications to which the mathematical programming models were originally applied, the methodology developed in this research may be easily extended to networks other than simply social networks. For example, multi-graphs capturing the interactions among individuals and Internet sites, layered networks, key components of layered infrastructure, and key "cities, regions, or tribes" within an influence structure can be investigated using this approach. Although the latter examples are not limited to non-cooperative networks, they do offer opportunities to influence the environments within which such networks thrive.

Assuming that data availability improves as analysis efforts progress, two methods to take advantage of information characterizing the nature of interpersonal ties are developed. This ultimately yields a new, theoretical approach measuring the strength of interpersonal relationships. This approach is similar to that offered by

Renfro [2001] and Clark [2005], but is not as data intensive as the former, and less reliant upon other sociometric methods than the latter. The first of two methods proposed in this area focuses primarily upon the relationships between similarity and tie strength, along the lines of homophily. The second method builds upon the conceptual relationships between multiplexity and tie strength [Granovetter, 1973]. The primary benefit of the similarity technique is that the multiplex data is most likely readily available within existing data. The primary detriment of the similarity technique is that all dimensions of a relationship are treated equally, an assumption that may not sufficiently represent reality. The second method obviates this assumption, although at the expense of increased analytic effort. The resulting decision theoretic model builds upon the sociological construct of the factors that may contribute to tie strength.

With this model of tie strength, different layers (or relationship dimensions) can be weighted, preferably from the perspective of the members within the target organization. Such measurements can and should be debated, since an inherent bias towards ones own world view is only human and the purpose of the approaches is to deal with networks that are intentionally obscuring their activities. Traditional sensitivity analysis techniques and new, dynamic weighting schemes are proposed to investigate these types of questions. Tie strength then serves as an input into classic network flow models as an arc capacity, representing the maximum amount of influence one actor may impose upon another [cf. Freeman et al., 1991; Renfro, 2001; Clark, 2005]. However, it is also proposed that tie strength is inversely proportional to the cost, or at least a component therein, of interpersonal communication. Either option offers a means to improve the network flow representation of social networks.

An additional improvement related to the application of network flow models to analyze social networks is the pair-wise valued measure of influence gains and losses. This measurement approach mathematically develops the relationships between an individual's personal characteristics and the power or persuasive capability

that results. Since a gain of influence could be garnered from a social position of power, the persuasive nature of the individual, or both, the overarching concept of charisma is considered. Logistic regression using individual characteristics and expert opinion regarding who is or is not persuasive relative to the others within the network yields a fitted logistic response value. This value can be interpreted as the estimated probability that an actor with certain characteristics is charismatic (i.e., can invoke influence via either their power or persuasive characteristics, or both). These values can then be compared on a pair-wise basis, for all pairs of individuals that are known to have at least one dimension of relationship. This ultimately yields a gain (or loss) multiplier which is incorporated into a generalized network flow formulation.

The value of tie-strength is essentially inferred from the multiplex relationships. The proposed methods to measure tie strength allow the aggregation of multiple social networks into a single, weighted network. Combining these techniques with the gains and losses concept, a logical extension of the network flow centrality measure proposed by Freeman et al. [1991], is offered. Although Freeman et al. merely speculated arc weights, this dissertation provides a means to determining a meaningful value, in addition to the methods developed by Renfro [2001] and Clark [2005]. The inclusion of individual-specific attributes, similar to the work by Clark [2005], potentially improves the fidelity of the social network model. Consequently, the inclusion of both topological and individual specific components of influence yield a generalized network flow centrality measure; this measure estimates the centrality of an individual based upon not only their connections, but the strength of their connections, as well as how their degree of influence is perceived by others.

The mathematical programming model of the generalized network flow social model may also be used to investigate potential courses of action levied against a non-cooperative network. As demonstrated in the case study in Chapter VII, the

efficacy of influence operations impinging upon susceptible individuals in order to influence inaccessible decision makers or opinion-leaders may be examined.

All of the methods presented in this dissertation include some form of sensitivity analysis. Although the general assumption of deterministic data is maintained in this research, it is clear that relationships, networks, and social roles are likely changing over time. Sensitivity analysis techniques offer a means to deal with this phenomenon in the interim. Snapshots in time, consistent with the deterministic assumptions of the approaches developed here, provide a means to represent dynamic change over time.

To supplement the theoretical contributions of measurement and methodological approaches as applied to layered social networks, practical contributions comprise the demonstration of the overall methodology in Chapter VIII and the associated suite of algorithms developed within the MATLAB environment. This documented code is available for immediate use, either for intelligence analysis or for future research efforts.

9.3 Recommendations for Future Research

Within the intersection of operations research and the sociological sciences, there remain a number of research opportunities. Refinements in existing social network analysis measures, such as those recommended in Appendix H regarding information centrality [Stephenson and Zelen, 1989], as well as the methods developed in this dissertation, are still needed.

Regarding the genre of network flow models, as discussed in Borgatti [2005] for example, different SNA measures are typically accompanied by underlying, often implicit, assumptions regarding the traversal mechanism of the entity of interest (influence, power, prestige, disease, rumors, etc.). Consequently, applying an inappropriate measure to a network phenomenon may give misleading or erroneous

results. It is suggested that applying certain mathematical programming techniques may also suffer this same weakness. Simply stated, different commodities of interpersonal interaction (communication of influence, rumors, packages, and so forth) have the potential to traverse a network in different ways, not all of which are amenable to the classic network flow problem. To bolster the research dealing with mappings between SNA and OR, particularly in the area of network flow models, the implications of flow typology should be considered in detail. Such a future analysis should include appropriate assumptions or improve network flow models via additional constraints. For example, for each possible combination of trajectories and method of transfer discussed in Borgatti [2005], a mapping between the network commodity, its behavior, and one of the following options could include: (1) a standard network flow formulation; (2) a network flow formulation with appropriate side constraints; (3) a network flow-like formulation; or (3) a recommendation for either another mathematical program, a simulation, or a heuristic.

Regarding the measure of interpersonal tie-strength based upon (topological) similarity, it may be of interest to investigate the theoretical and practical feasibility of the application of 19 other similarity measures as described in Yin and Yasuda [2005]. Simple, robust, and accurate measures of tie strength that are quick to calculate and grounded in sociological theory and intelligence needs are paramount to investigating these types of networks. Again, this method is advantageous in that it is predicated upon discovered layers of relationships; therefore, the need for a cross-cultural methodology is not a primary concern as existing connections themselves contain the elements important to that organization. Of course, determining what those exact elements are may be challenging based upon the nature of the data collected. Unfortunately, most, if not all, current similarity approaches assume each layer is equally important with regard to its contribution to tie strength.

Since one of the objectives of this research includes disrupting the operation or efficiency of a target network, accounting for negative ties may also be of interest.

How to explain, model, and exploit negative ties is critical, because destroying these networks will inevitably involve some element of turning the members against one another. The cohesion research offered by Downs [2006] may prove a useful start in finding methods to disrupt networks via negative ties, and measuring their efficacy.

The network disruption objective also suggests that the key player (KPP-1) problem is of interest. Recall that KPP-1 finds a subset of actors within a network that, if removed, will maximally disrupt communication among the remaining actors. To date, no mathematical program equivalent to the heuristic procedure (and objective function) has been successfully developed. However, such a program would benefit the analysis process in ways similar to those explored for KPP-2. For example, disruption could be tailored to meet specific decision maker requirements, as opposed to adhering strictly to the objective function defined by Borgatti [2003a]. Additionally, multiple optimal solutions, if they exist, would offer alternative plans and insights into the structural roles of individuals within the context of KPP-1. Finally, the exploration and development of heuristic techniques to address both KPP problems is of interest due to the computational complexity involved. This would facilitate KPP solutions for larger networks, either social, physical, or any network-like system.

Visualization and data continue to be a limiting factor when exploring sociometric methods. Data, particularly unclassified data for use in open, academic research, is often difficult to obtain. Although ongoing efforts to improve both areas exist, there does not appear to be any visualization techniques or software applications that lend themselves specifically to viewing layered networks. Data generation of networks exhibiting properties similar to those of the target networks may be accomplished using the Organizational Risk Analyzer tool Carley [2006] or by extending the initial network generation work of Sterling [2004]. One possible means to visualize layered networks could start with plotting the entire, aggregated network as a single layer, with the initial layout determined by the analyst. Assuming the

network nodes are at a given (x, y) location and the network resides within the plane $z = 0$, any subsequent layers l will reside within the planes $z = l, l \in \mathbb{Z}^+$, with each node using the same corresponding (x, y) coordinate in $z = 0$.

A host of challenges remain. Social network modeling will remain a fertile area of research for the the operations researcher.

9.3.1 Conclusion

> The war on terrorism is going to be a long one; it is "...both a battle of arms and a battle of ideas–a fight against terrorist networks and against their murderous ideology" [DOD, 2006b, pg. v,22].

Considering the nature of this war, understanding the enemy is paramount. No longer fitting the traditional paradigm of combat between great armies, this war involves not only defeating the individuals actively threatening our National Security, but also mitigating the environments that nurture the development and continuity of such groups. In order to accomplish this, the appropriate communities must at a minimum (1) improve the understanding of why people would undertake such activities; (2) identify limitations or vulnerabilities existing within non-cooperative networks and how to exploit them; and, (3) determine what repercussions may follow an operation to minimize the likelihood that actions executed inadvertently contribute to the environments that promote extremism.

Appendix A. Code: rbap.m

This code performs the reach-based assessment of position (RBAP) measure developed in Chapter IV.

```
function [rbapR] = rbap(X, alpha)
% Usage rbap(X, alpha)
% This is the fast(er) version
% - X is the sociomatrix (adjacency matrix) of a network (graph);
%   alpha is an attenuation factor, between 0 and 1, that is
%   specified by the user to indicate the importance or expected
%   degradation of influence associated with longer-distance
%   paths between individuals.  An alpha = 0 causes this measure
%   to revert to simple out-degree centrality.
%   An alpha = 1 causes this measure to be bounded above by the
%   number of other actors that can be reached/influenced by a
%   given actor.
%
%   This procedure provides the reach centrality/influence
%   measure developed by Hamill, Deckro, Chrissis, and Mills
%
% - X may be a symmetric (asymmetric) graph (digraph).
% - The output (rbapR) is a non-normalized index of each actor's
%   influence through reachability (or radiality), based upon
%   the number (and distance) of shortest paths to all other
%   actors.
%
% - jth / last modified on 8 FEB 06
   tic
%   Get dimensions of input
    [n, m] = size(X);
%   Reachability matrix is used to track shortest paths between
%   actors
%   First step, reach of one, is derived from the input (X)
    rmatrix = X;
%   vector used to zero out diagonals
    zdiag = [1:(n+1):(n*n)]';
%   initialize reach normalization matrix
%   number of other actors reached... at step p
    nAR = sum(X,2);
%   first element in power sum is the adjacency matrix
```

```
    tRBAP = X;
%   temporary power matrix used to speed up calculations
    txPow = X;
%   for shortest paths of length 2 to the maximum (n-1)
    for i = 2:(n-1);
%       find non-zero entries in previous power of X
        lastStep = find(rmatrix==0);
%       raise X to the next power and zero out diagonals
        txPow = txPow*X;
        xPow = txPow;
        xPow(zdiag) = 0;
%       find non-zero entries in the next power of X
        nextStep = find((xPow)>0);
%       find entries that become non-zero for the first time
        reach = intersect(nextStep, lastStep);
%       update the reachability matrix
        rmatrix(reach) = i;
%       initialize a temporary matrix of zeros to capture
%       number of paths
        tmpR = zeros(n,n);
%       temporary matrix captures the number of (shortest)
%       paths of length i
        tmpR(reach) = xPow(reach);
%       can break out of program early if this is all zeros...
        if nnz(tmpR)==0
            break
        end
%       need to divide next iterate by product of nAR
%       (if it is not zero)
        for k = 1:n
            if nAR(k,1)>0
                tmpR(k,:) = tmpR(k,:)/nAR(k);
            end
        end %k
%       update measure results
        tRBAP = tRBAP + (alpha^(i-1))*tmpR;
%       update for next iteration the temporary matrix and
%       divisor captures the number of new actors reached
        tmpAR = zeros(n,n);
        tmpAR(reach) = 1;
        tmpARdiv = sum(tmpAR,2);
        nAR = nAR.*tmpARdiv;
```

```
    end
%   Find row sum for each actor over all other actors, over all
%   sp lengths.  Note that path lengths have been attenuated
%   already
    rbapR = sum(tRBAP,2);
   toc
```

Appendix B. Code: rbapsa.m

This code performs the sensitivity analysis of the attenuation factor α within the reach-based assessment of position (RBAP) measure.

```
function [pdata] = rbapsa(X)
% Usage rbapsa(X)
%
% - X is the sociomatrix (adjacency matrix) of a network (graph);
% - The attenuation factor, alpha, is set to values in [0, 1] and
%   the RBAP measure is calculated for each.
% - The results are then plotted to facilitate sensitivity analysis
%   of alpha, and its impact upon the ranking between individuals
%   based upon the assumption of how much (or little) indirect
%   communication is distance-attenuated.
%   Last revised 12 JAN 06 - jth
%   ***** Begin Function *****
    [n, m] = size(X);   % get dimensions of input
    a = 0:0.1:1;       % set range of alpha
    [ra, ca] = size(a); % get dimensions of output
%   for each setting of alpha, call RBAP and store results
    for i = 1:ca
        rbc = rbap(X,a(i));
        pdata(:,i) = rbc;
    end
%   Plot results
    plot(a, pdata);
```

Appendix C. Code: KPP-2 (NR.m)

This code finds the minimum size *kp*-set required to reach all nodes, using m-steps or less between a key player and his assigned actor. Note that this and other programs also use the function `Mreach.m` presented at the end of this appendix.

```
function [ kp_nr ] = NR( X, MR, enum )
% Useage NR(X, MR, enum)
%      X - Adjacency matrix
%     MR - Maximum reach allowed by any key player
%          (-1 for maximum (N-1))
%   enum - Enumerate solutions until infeasible or maximum number have
%          been found (maxsol is currently set at 100 solutions)
%   Updated 4 JUL 2006, jth
%   Requires the MATLAB Optimization Toolbox
%   The current objective function assumes that all options are equal.
%   This, however, is amenable to change to accommodate external and
%   internal costs associated with selecting/targeting the individual
%   and the workload demanded/assumed of the individual, respectively.
%
%    Notes:
%    -- This program appears to be computationally competitive to the
%    heuristic provided by Borgatti.  Large-scale problems (n > 1000),
%    however, may require linking up CPLEX to MATLAB, or using other
%    specialized optimization programs with more efficient algorithms
%    (e.g. LINGO.
    tic
%    ***** initialize variables *****
    maxSol = 100;       % maximum number of solutions (for enum = 1)
    numSol = 0;         % current number of solutions
    N = size(X,1);      % dimension of X
    A = zeros(N);       % constraint matrix
    b = - ones(N,1);    % column vector as RHS
    f = ones(N,1);      % column vector as obj. function coeff.
    kp_nr = [];         % storage for solutions
    firstSol = 1;       % boolean flag to control enumeration loops
    foundall = 0;       % boolean flag to control enumeration loops
    kpK = 0;            % kp-set size, facilitates enumeration loops
%    Develop constraint matrix based upon inputs X and reach MR
%    Note:  See Mreach code for details
```

```
    if MR == -1
        rX = Mreach(X,(N-1));   % calculate maximum reach matrix
    else
        rX = Mreach(X,MR);      % calculate reach matrix up to MR
    end
    dc = find(rX>0);    % find non-zeros in reach matrix
    A(dc) = - 1;        % set up constraint matrix, A
    if enum
        while not(foundall)
            % column vector of x as solution
            [mpSol, obj, flag] = bintprog(f, A, b);
            if firstSol
                kpK = obj;      % size of min. dom. set
                firstSol = 0;
            end % if firstSol...
            if flag == -2 | (numSol == maxSol) | (obj ~= kpK)
                foundall = 1
            else
                kp_set = find(mpSol(1:N)==1);
                kp_nr = [kp_nr kp_set];
                b = [b; (kpK - 1)];      % add extra rhs
                A = [A; mpSol'];         % add extra constraint
                numSol = numSol + 1;
            end % if flag...
        end % while not...
%       Plot data in histogram
%       (only required when enumerating solutions)
        [a, b] = size(kp_nr);
        pdata = [];
        xaxis = 1:N;
        for j = 1:a
            pdata = [pdata kp_nr(j,:)];
        end
        hist(pdata,xaxis);
        xlim([0 N]);
        ylim([0 b]);
    else
%       column vector of x as solution
        mpSol = bintprog(f, A, b);
        kp_nr = find(mpSol(1:N)==1) % key players
    end % if enum...
toc     % time elapsed
```

The following function is used within some of the KPP-2 optimization problems, and may also be used stand-alone.

```
function [rmatrix] = Mreach( X, M )
% Useage Mreach(X,M)
% X - Sociomatrix
% M - Maximum number of steps allowed for reachability.
% This function determines which actors are reachable from others,
% based upon a limit of M-steps or less.
% This currently assumes symmetric relationships.
[n, m] = size(X); rmatrix = X; for i = 2:M
    lastStep = find(rmatrix==0);
    nextStep = find((X^i)>0);
    reach = intersect(nextStep, lastStep);
    rmatrix(reach) = i;
    if (sum(find(rmatrix==0))==0)   % all
        break
    end
end for j = 1:n
    rmatrix(j,j)=1;
end
```

Appendix D. Code: KPP-2 (FNRK)

This code minimizes the number of actors missed if the size of the key player set is less than the m-domatic number of the graph. For this code, the maximum reach allowed m is assumed to be either 1 or 2, which is modeled by `FNRK1.m` and `FNRK2.m`, respectively. The output is a listing of the key players, a separator of 0, and the total number of actors missed.

```
function [ kp_fnrk1 ] = FNRK1( X, K, enum )
%Usage FNRK1( X, K, enum )
%     X - Adjacency matrix
%     K - Size of kp-set
%  enum - Enumerate solutions until infeasible or maximum
%         number has been found
%         (maxsol is currently set at 300 solutions)
%   Updated 22 MAR 2006, jth
%   Requires the MATLAB Optimization Toolbox
%   Seeks to minimize f(x) : Ax >= 1, x is binary {0, 1}
%   Objective is FNRK1
    tic
    maxSol = 300;
    numSol = 0;
    N = size(X,1);
    kp_fnrk1 = [];
    R1 = (X' + eye(N));
    A = [(-R1) (-eye(N))];
    Aeq = [ones(1,N) zeros(1,N)];
    b = - ones(N,1);
    beq = K;
%   column vector as obj. function coeff.
    f = [zeros(N,1); ones(N,1)];
%   While continue, keep solving until the cardinality of the kp set
%   increases by one...
    if enum
        foundall = 0;
        kpK = 999999;
        numMiss = 999999;
        while not(foundall)
%           column vector of x as solution
```

```
            [mpSol, obj, flag] = bintprog(f, A, b, Aeq, beq);
%           key players
            tmp_kp_fnrk1 = find(mpSol(1:N)==1);
%           missed players
            tmp_miss_fnrk1 = nnz(mpSol((N+1):2*N));
            if flag == -2 | (numSol > maxSol)
                foundall = 1
            else
                if (nnz(tmp_kp_fnrk1) <= kpK) &
                       (tmp_miss_fnrk1 <= numMiss)
                    numSol = numSol + 1;
                    kp_fnrk1 =
                      [kp_fnrk1 [tmp_kp_fnrk1; 0; tmp_miss_fnrk1]];
                    kpK = nnz(tmp_kp_fnrk1);
                    numMiss = tmp_miss_fnrk1;
%                   add extra rhs for additional constraint
                    b = [b; (kpK - 1)];
                    A = [A; [(mpSol(1:N))' zeros(1,N)] ];
                else
                    foundall = 1;
                end
            end
        end
    else
%       column vector of x as solution
        mpSol = bintprog(f, A, b, Aeq, beq);
%       key players
        tmp_kp_fnrk1 = find(mpSol(1:N)==1);
%       missed players
        tmp_miss_fnrk1 = nnz(mpSol((N+1):2*N));
        kp_fnrk1 = [tmp_kp_fnrk1; 0; tmp_miss_fnrk1];
    end
toc

function [ kp_fnrk2 ] = FNRK2( X, K, enum )
%Usage FNRK2( X, K, enum )
%     X - Adjacency matrix
%     K - Size of kp-set
%  enum - Enumerate solutions until infeasible or maximum
%         number has been found
%         (maxsol is currently set at 300 solutions)
%   Updated 22 MAR 2006, jth
%   Requires the MATLAB Optimization Toolbox
```

```
%   Seeks to minimize f(x) : Ax >= 1, x is binary {0, 1}
%   Objective is FNRK2
    tic
    maxSol = 300;
    numSol = 0;
    N = size(X,1);
    kp_fnrk2 = [];
    R1 = (X' + eye(N));
    tR2 = (X')^2 + R1;
    dc = find(tR2>0);
    R2 = zeros(N);
    R2(dc) = - 1;
    A = [(-R1) R2 (-eye(N))];
    Aeq = [ones(1,2*N) zeros(1,N)];
    b = - ones(N,1);
    beq = K;            % column vector as RHS
    b = - ones(N,1);    % column vector as RHS
%   column vector as obj. function coeff.
    f = [zeros(2*N,1); ones(N,1)];
%   While continue, keep solving until the cardinality of the kp set
%   increases by one...
    if enum
        foundall = 0;
        kpK = 999999;
        numMiss = 999999;
        while not(foundall)
%           column vector of x as solution
            [mpSol, obj, flag] = bintprog(f, A, b, Aeq, beq);
%           key players, 1-step
            tmp1 = mpSol(1:N);
%           key players, 2-step
            tmp2 = mpSol((N+1):2*N);
%           all key players
            tmp_kp_fnrk2 = [ find(tmp1==1); find(tmp2==1) ];
%           missed players
            tmp_miss_fnrk2 = nnz(mpSol((2*N+1):3*N));
            if flag == -2 | (numSol > maxSol)
                foundall = 1
            else
                if (nnz(tmp_kp_fnrk2) <= kpK) &
                   (tmp_miss_fnrk2 <= numMiss)
                    numSol = numSol + 1;
```

```
                    kp_fnrk2 =
                      [kp_fnrk2 [tmp_kp_fnrk2; 0; tmp_miss_fnrk2]];
                    kpK = nnz(tmp_kp_fnrk2);
                    numMiss = tmp_miss_fnrk2;
%                   add extra rhs for additional constraint
                    b = [b; (kpK - 1)];
                    A = [A; [(mpSol(1:2*N))' zeros(1,N)] ];
                else
                    foundall = 1;
                end
            end
        end
    else
%       column vector of x as solution
        mpSol = bintprog(f, A, b, Aeq, beq);
%       key players, 1-step
        tmp1 = mpSol(1:N);
%       key players, 2-step
        tmp2 = mpSol((N+1):2*N);
%       all key players
        tmp_kp_fnrk2 = [ find(tmp1==1); find(tmp2==1) ];
%       missed players
        tmp_miss_fnrk2 = nnz(mpSol((2*N+1):3*N));
        kp_fnrk2 = [tmp_kp_fnrk2; 0; tmp_miss_fnrk2];
    end
toc     % time elapsed
```

Appendix E. Code: KPP-2 (FNR.m)

This code minimizes the number of key players required to cover (1-PC) percent of the network members, with the additional constraint upon the maximum reach allowed between a key player and its assigned actor(s). The output is a listing of the key players, a separator of 0, and the total number of actors covered.

```
function [ kp_fnrm ] = FNR( X, MR, PC, enum )
% Useage FNR(X, MR, PC, enum)
%      X - Adjacency matrix
%     MR - Maximum reach allowed by any key player
%          (-1 for maximum (N-1))
%     PC - Percent of network that can be MISSED
%   enum - Enumerate solutions until infeasible or maximum number
%          have been found (1=Yes, 0=No)
%          (maxsol is currently set at 100 solutions)
%   Updated 4 JUL 2006, jth
%   Requires the MATLAB Optimization Toolbox
%   Seeks to minimize f(x) : Ax >= 1, x is binary {0, 1}
%   Objective is FNRm
%   ***** initialize variables *****
    maxSol = 100;       % maximum number of solutions (for enum = 1)
    numSol = 0;         % current number of solutions
    N = size(X,1);      % dimension of X
    U = floor(PC*N);
    A = zeros(N);       % constraint matrix
    b = - ones(N,1);    % column vector as RHS
    f = ones(N,1);      % column vector as obj. function coeff.
    kp_fnrm = [];       % storage for solutions
    firstSol = 1;       % boolean flag to control enumeration loops
    foundall = 0;       % boolean flag to control enumeration loops
    kpK = 0;            % kp-set size, facilitates enumeration loops
%   Develop constraint matrix based upon inputs X and reach MR
%   Note:  See Mreach code for details
    if MR == -1
        rX = Mreach(X,(N-1));   % calculate maximum reach matrix
    else
        rX = Mreach(X,MR);      % calculate reach matrix up to MR
    end
    dc = find(rX>0);    % find non-zeros in reach matrix
```

```
    A(dc) = - 1;        % set up constraint matrix, A
    A = [A -eye(N); zeros(1,N) ones(1,N)];
    b = [-ones(N,1); U];            % column vector as RHS
%   column vector as obj. function coeff.
    f = [ones(1,N) zeros(1,N)]';
    tic
%   While continue, keep solving until the cardinality of the kp set
%   increases by one...
    if enum
        while not(foundall)
%           column vector of x as solution
            [mpSol, obj, flag] = bintprog(f, A, b);
            if firstSol
                kpK = obj;      % size of min. partially dom. set
                firstSol = 0;
            end % if firstSol...
            if flag == -2 | (numSol == maxSol) | (obj ~= kpK)
                foundall = 1
            else
                kp_set = find(mpSol(1:N)==1);
                missed = sum(mpSol((N+1):(2*N)));
                kp_fnrm = [kp_fnrm [kp_set; 0; missed]];
%               add extra rhs
                b = [b; (kpK - 1)];
%               add extra constraint
                A = [A; [mpSol(1:N)' zeros(1,N)]];
                numSol = numSol + 1;
            end  % if flag...
        end  % while...
    else
        mpSol = bintprog(f, A, b);  % column vector of x as solution
%       key players
        kp_fnrm = [find(mpSol(1:N)==1); 0; sum(mpSol((N+1:2*N)))]
    end % if enum...
toc     % time elapsed
```

Appendix F. Code: KPP-2 (DNR.m)

This code finds the minimum size *kp*-set required to reach all nodes, using m-steps or less between a key player and his assigned actor, while distributing the workload among the key players. This is accomplished by using the inverse of the out-degree of each individual as the cost coefficient in the objective function.

```
function [ kp_dnr, f ] = DNR( X, MR, enum )
% Useage DNR(X, MR, enum)
%     X - Adjacency matrix
%    MR - Maximum reach allowed by any key player
%         (-1 for maximum (N-1))
%  enum - Enumerate solutions until infeasible or maximum
%         number have been found
%         (maxsol is currently set at 100 solutions)
%  Updated 5 JUL 2006, jth
%  Requires the MATLAB Optimization Toolbox
%   ***** initialize variables *****
    maxSol = 100;       % maximum number of solutions (for enum = 1)
    numSol = 0;         % current number of solutions
    N = size(X,1);      % dimension of X
    A = zeros(N);       % constraint matrix
    C = zeros(N);       % temporary cost matrix
    b = - ones(N,1);    % column vector as RHS
    kp_dnr = [];         % storage for solutions
    firstSol = 1;       % boolean flag to control enumeration loops
    foundall = 0;       % boolean flag to control enumeration loops
    kpK = 0;            % kp-set size, facilitates enumeration loops
%   Develop constraint matrix based upon inputs X and reach MR
%   Note:  See Mreach code for details
    if MR == -1
        rX = Mreach(X,(N-1));   % calculate maximum reach matrix
    else
        rX = Mreach(X,MR);      % calculate reach matrix up to MR
    end
    dc = find(rX>0);    % find non-zeros in reach matrix
    A(dc) = - 1;        % set up constraint matrix, A
    C(dc) = 1;          % set up cost matrix to determine f
    f = 1./sum(C,2);
    tic
```

```
    if enum
        while not(foundall)
%           column vector of x as solution
            [mpSol, obj, flag] = bintprog(f, A, b);
            if firstSol
                kpK = obj;      % size of min. dom. set
                firstSol = 0;
            end % if firstSol...
            if flag == -2 | (numSol == maxSol) | (obj ~= kpK)
                foundall = 1
            else
                kp_set = find(mpSol(1:N)==1);
                kp_dnr = [kp_dnr kp_set];
                b = [b; (kpK - 1)];  % add extra rhs
                A = [A; mpSol'];     % add extra constraint
                numSol = numSol + 1;
            end % if flag...
        end % while not...
%       Plot data in histogram
        [a, b] = size(kp_dnr);
        pdata = [];
        xaxis = 1:N;
        for j = 1:a
            pdata = [pdata kp_dnr(j,:)];
        end
        hist(pdata,xaxis);
        xlim([0 N]);
        ylim([0 b]);
    else
        mpSol = bintprog(f, A, b);  % column vector of x as solution
        kp_dnr = find(mpSol(1:N)==1) % key players
    end % if enum...
toc     % time elapsed
```

Appendix G. Code: KPP-2 (PMED)

This appendix outlines the process to apply the p-Median problem to KPP-2. Initial efforts revealed that this class of problem is too computationally intensive for the branch-and-bound method implemented within the MATLAB optimization toolkit environment.

Consequently, more efficient solution algorithms are recommended in order to implement this type of analysis. For the following process proposed, several files are required. The first sample file depicts the mathematical model developed in LINGO. This model requires network data as input. As an illustrative example, the input for the ('methodscamp') network accompanying the key player software is also provided.

```
MODEL: ! P-Median approach to KPP2;

! Sets defining the data structure;
SETS:
    KP     / @FILE('pmed_mc.LDT') /: FCOST, KPSET;
    ACTORS / @FILE('pmed_mc.LDT') /: ;
    ARCS( KP, ACTORS )               : COST, X;
ENDSETS

DATA: ! The cost (i,j) to reach actor j from kp i;
  COST = @FILE('pmed_mc.LDT');
! The maximum kp-set size;
  KPMAX = 2;
! Fixed cost of coopting a key player may be included;

! FCOST = (csv row vector of cost/actor here) ;

ENDDATA

! The objective is to Minimze the weighted distance from ! key
players to all actors;
  [TTL_COST] MIN = @SUM( ARCS | (COST #GT# 0) :
                  -1/COST * X);
! + @SUM( KP: FCOST * KPSET);
```

```
! Assign all actors to at least one key player;
  @FOR( ACTORS ( J): [ASSIGN]
  @SUM( KP( I) | (COST(I, J) #GT# 0 #OR# I #EQ# J):
        X(I, J)) = 1
      );

! Actors cannot be assigned to someone not in the KP SET;
  @FOR( KP( I):
  @FOR( ACTORS ( J) | (COST(I, J) #GT# 0 #OR# I #EQ# J) :
        [KPA] X( I, J) <= KPSET( I) );
      );

! Number of key players allowed in the KP SET;
  [KPTOT] @SUM( KP( I): KPSET( I)) = KPMAX;

! Make KPSET choice (OPEN) binary(0/1);
  @FOR( KP: @BIN( KPSET));

END
```

The following sample file is the input corresponding to the ('methodscamp') network data used as a demonstration with the key player software developed by Borgatti [2003b]. Given a network of size N, the indices should range from K1 ...KN, and A1 ...AN, respectively. For this example, $N = 18$. The reach matrix is as expected, with the exception that the reachability values must be obtained using the transpose of the sociomatrix. Note that the function Mreach.m provided in Appendix C may be used to determine these values. Since the objective function uses the additive inverse of these values, this matrix could be constructed in order to place restrictions upon maximum allowable reach, where the value would be 0 if two actors could not reach each other in a given number of steps. In addition, although the values here are based upon a dichotomous network, other measures of distance between actors may be used.

```
!  P-Median data for Methodscamp data;
!  Key player indices;
   K1..K18 ~
!  Actor indices;
```

```
   A1..A18 ~
!  Reach matrix;
   1,4,2,1,1,2,2,2,1,2,4,1,3,1,2,3,4,3,
   4,1,5,5,5,6,4,5,3,4,1,4,3,4,2,1,1,2,
   2,5,1,1,1,2,1,2,3,4,5,3,2,3,3,4,4,3,
   1,5,1,1,2,1,1,1,2,3,5,2,2,2,3,4,4,3,
   1,5,1,2,1,1,1,2,2,3,5,2,2,2,3,4,4,3,
   2,6,2,1,1,1,2,1,3,4,6,3,3,3,4,5,5,4,
   2,4,1,1,1,2,1,1,3,4,4,3,1,3,2,3,3,2,
   2,5,2,1,2,1,1,1,3,4,5,3,2,3,3,4,4,3,
   1,3,3,2,2,3,3,3,1,1,3,1,2,1,1,2,3,2,
   2,4,4,3,3,4,4,4,1,1,4,1,3,1,2,3,4,3,
   4,1,5,5,5,6,4,5,3,4,1,4,3,4,2,1,1,2,
   1,4,3,2,2,3,3,3,1,1,4,1,3,1,2,3,4,3,
   3,3,2,2,2,3,1,2,2,3,3,3,1,3,1,2,2,1,
   1,4,3,2,2,3,3,3,1,1,4,1,3,1,2,3,4,3,
   2,2,3,3,3,4,2,3,1,2,2,2,1,2,1,1,2,1,
   3,1,4,4,4,5,3,4,2,3,1,3,2,3,1,1,1,1,
   4,1,4,4,4,5,3,4,3,4,1,4,2,4,2,1,1,1,
   3,2,3,3,3,4,2,3,2,3,2,3,1,3,1,1,1,1
```

The solution to this mathematical program is KPSET(K1) = KPSET(K16) = 1, indicating that actors 1 and 16 are the key players. The objective function value of -15 is equivalent to the normalized objective function for Borgatti's heuristic output of $15/18 = 0.8333$.

Appendix H. Note on Information Centrality

This note addresses an underlying, potential mathematical issue with the information centrality measure developed by Stephenson and Zelen [1989]. In the current effort to develop new means to study network structures, a MATLAB program to enumerate all paths within a graph or digraph was developed (See Appendix J). Due to interest and previous applications [e.g., Clark, 2005] of the information centrality measure, the example graph used by Stephenson and Zelen was used to validate the path enumeration code. Interestingly, Stephenson and Zelen missed two paths. The graph and the path lists are shown below. The paths highlighted, 3-4-1-2-5 and 4-1-2-5, are those either missed or ignored by Stephenson and Zelen.

As the information centrality measure is "based on the information contained in *all possible paths* between pairs of points," (emphasis added) the next logical step is to compare results, using their methodology, between datasets. This initial effort is investigating dichotomous graphs only; applying this methodology to valued graphs is also suspect.

As defined in their article, the information in the combined path $\mathbf{I}_{ij}$ is the sum of all the elements of the $\mathbf{D}_{ij}^{-1}$ matrix [Stephenson and Zelen, 1989, pg. 9-10]. Given all k paths between i and j, the $\mathbf{D}_{ij}$ matrix is $k \times k$, and defined as

Table 8.1: Paths for each node pairs

Node Pair	Corresponding Paths
1-2	1-2, 1-5-2, 1-4-3-2
1-3	1-4-3, 1-2-3, 1-5-2-3
1-4	1-4, 1-2-3-4, 1-5-2-3-4
1-5	1-5, 1-2-5, 1-4-3-2-5
2-3	2-3, 2-1-4-3, 2-5-1-4-3
2-4	2-1-4, 2-3-4, 2-5-1-4
2-5	2-5, 2-1-5, 2-3-4-1-5
3-4	3-4, 3-2-1-4, 3-2-5-1-4
3-5	3-2-5, 3-2-1-5, 3-4-1-5, ***3-4-1-2-5***
4-5	4-1-5, 4-3-2-5, 4-3-2-1-5, ***4-1-2-5***

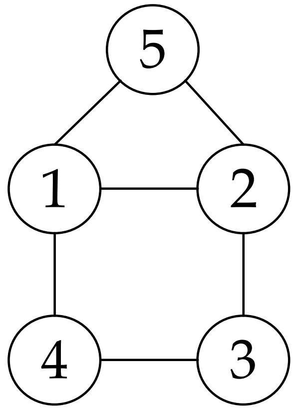


Figure 8.1: Exemplar Network [Stephenson and Zelen, 1989]

$\mathbf{D}_{ij}(r,s) =$ number of lines in paths$, r = s$

$\mathbf{D}_{ij}(r,s) =$ number of lines in common between path r and path $s, r \neq s$

As an example, node pair 1-2 has 3 paths, none of which have any arcs in common; the resulting calculations are given by $\mathbf{D}_{12}$.

$$\mathbf{D}_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \Rightarrow \mathbf{D}_{12}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.333 \end{bmatrix} \Rightarrow I_{12} = 1.833$$

Additionally, node pair 1-4 has 3 paths, two of which (the second and third path) have 2 arcs in common; the resulting calculations are shown by $\mathbf{D}_{14}$.

$$\mathbf{D}_{14} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 2 \\ 0 & 2 & 4 \end{bmatrix} \Rightarrow \mathbf{D}_{14}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & -0.25 \\ 0 & -0.25 & 0.375 \end{bmatrix} \Rightarrow I_{14} = 1.375$$

The results of both node pairs with missing paths are consistent with the data presented by the authors. For example, ignoring the missed path for node pair 3-5,

the calculations for $\mathbf{D}_{35}$ match those described in the article.

$$\mathbf{D}_{35} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 1 \\ 0 & 1 & 3 \end{bmatrix} \Rightarrow \mathbf{D}_{35}^{-1} = \begin{bmatrix} 0.6154 & -0.2308 & 0.0769 \\ -0.2308 & 0.4615 & -0.1538 \\ 0.0769 & -0.1538 & 0.3846 \end{bmatrix} \Rightarrow I_{35} = 0.8462$$

Note that these calculations, using the construct developed by Stephenson and Zelen, match those presented on Stephenson and Zelen [1989, pg. 10]. More importantly, these calculations, and all others, also coincide to the results when using the overall procedure as described on Stephenson and Zelen [1989, pg. 12]. This procedure, taken verbatim from Stephenson and Zelen [1989], is presented as follows.

Consider a network with n points where every pair of points is reachable. Define the $n \times n$ matrix $\mathbf{B} = (b_{ij})$ by:

$$b_{ij} = \begin{cases} 0 \text{ if points } i \text{ and } j \text{ are incident} \\ 1 \text{ otherwise} \end{cases}$$

$$b_{ii} = 1 + \text{degree of point } (i)$$

Defining $\mathbf{C} = (c_{ij}) = \mathbf{B}^{-1}$, $\mathbf{I}_{ij} = (c_{ii} + c_{jj} - 2c_{ij})^{-1}$ and using this information, revisiting the earlier calculations, yields the following.

$$\mathbf{B} = \begin{bmatrix} 4 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 3 & 0 & 1 \\ 0 & 1 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 \end{bmatrix}$$

$$\Rightarrow \mathbf{C} = \begin{bmatrix} 0.2764 & 0.0036 & -0.1055 & -0.0145 & 0.0400 \\ 0.0036 & 0.2764 & -0.0145 & -0.1055 & 0.0400 \\ -0.1055 & -0.0145 & 0.4218 & 0.0582 & -0.1600 \\ -0.0145 & -0.1055 & 0.0582 & 0.4218 & -0.1600 \\ 0.0400 & 0.0400 & -0.1600 & -0.1600 & 0.4400 \end{bmatrix}$$

$$I_{12} = (c_{11} + c_{22} - 2c_{12})^{-1} = (0.2764 + 0.2764 - 2 \cdot 0.0036)^{-1} = 1.8328 \approx 1.8333$$

$$I_{14} = (c_{11} + c_{44} - 2c_{14})^{-1} = (0.2764 + 0.4218 - 2 \cdot (-0.0145))^{-1} = 1.3751 \approx 1.3750$$

$$I_{35} = (c_{33} + c_{55} - 2c_{35})^{-1} = (0.4218 + 0.4400 - 2 \cdot (-0.1600))^{-1} = 0.8461 \approx 0.8462$$

These comparisons confirm that the steps used to develop the methodology and the actual calculations taken to assess the measure are identical, within round-off error. This provides satisfactory confirmation that the analytic approach is consistent with the data presented thus far. However, the fact that paths were missed causes some concern for the validity of the measure, at least in its claim to account for the information 'along all paths.' Differences in results can, at this point in the review, only be ascertained when doing the calculations 'by hand,' as the overall methodology (i.e., the construction of the **C** matrix) currently allows no room to account for this oversight.

Revisiting the information content for node pairs 3-5 and 4-5 via calculations 'by hand,' the actual values–assuming their approach is correct–should be as follows.

$$\mathbf{D}_{35} = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 3 & 1 \\ 1 & 0 & 1 & 4 \end{bmatrix} \Rightarrow \mathbf{D}_{35}^{-1} = \begin{bmatrix} 0.7838 & -0.3243 & 0.1892 & -0.2432 \\ -0.3243 & 0.5135 & -0.2162 & 0.1351 \\ 0.1892 & -0.2162 & 0.4595 & -0.1622 \\ -0.2432 & 0.1351 & -0.1622 & 0.3514 \end{bmatrix}$$

$$\Rightarrow I_{35} = 0.8649 \neq 0.8462$$

$$\mathbf{D}_{45} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 1 & 1 & 4 & 0 \\ 1 & 1 & 0 & 3 \end{bmatrix} \Rightarrow \mathbf{D}_{45}^{-1} = \begin{bmatrix} 0.7838 & 0.1892 & -0.2432 & -0.3243 \\ 0.1892 & 0.4595 & -0.1622 & -0.2162 \\ -0.2432 & -0.1622 & 0.3514 & 0.1351 \\ -0.3243 & -0.2162 & 0.1351 & 0.5135 \end{bmatrix}$$

$$\Rightarrow I_{45} = 0.8649 \neq 0.8462$$

Some interesting observations are readily apparent. First, each node pair has three associated paths. Second, both node-pairs that are missing paths (3-5 and 4-5) have identical values for both the initial calculation and the actual value suggested by Stephenson and Zelen–differing by a constant of ≈ 0.0187. Although this error is seemingly small, it is an indication that the underlying methodology is flawed and the interpretation of the measure as advertised is suspect. Applying this measure to larger and/or valued graphs is likely to further confound or mislead the analyst.

Based upon the test network described in the article, UCINET (Version 6) and NetMiner II currently calculate the information centrality measure exactly as described by Stephenson and Zelen [1989]. Due to the prevalence of sociological studies applying the information centrality measure, informing the academic community of this issue, preferably with a corresponding resolution, would provide immediate value.

Appendix I. Code: jaccard.m

This function calculates the Jaccard similarity coefficient for a given network comprised of one or more layers. To deal with multi-layered networks, the node-edge adjacency matrices for each layer must be horizontally concatenated. The output is a symmetric matrix, $\mathbf{S}_{JS}$, where each (i, j)th element represents the strength of the tie between actors i and j.

```
function [js] = jaccard(NE)
% Usage jaccard(X)
%
% - NE  is the node-edge adjacency matrix of one or more layers of
%   a social network.  This automatically implies network symmetry.
% - js is the tie strength based upon the Jaccard Similarity
%   Coefficient;
%   This value ranges from zero to one
%
% - jth / last modified on 2 AUG 06
    [N, numE] = size(NE);
    js = zeros(N);
    for i = 1:(N-1)
        for j = (i+1):N
            js(i,j) = sum(NE(i,:)& NE(j,:)) / sum(NE(i,:)|NE(j,:));
            js(j,i) = js(i,j);
        end % for j...
    end % for i...
```

Appendix J. Code: enumeratePaths.m

This MATLAB function enumerates all paths within a graph or digraph. Depending upon the size and complexity of the graph, this function can be very expensive computationally. The input is a sociomatrix. The output is a multi-dimensional cell array with the ith entry containing all paths of length $i + 1$. The code below currently limits the enumeration procedure to paths of length 6 or less, but could be modified (see below) to continue until all paths up to length $(n - 1)$ are found. As this function was developed primarily for the testing and investigation into the nature of small social networks, this function has a great deal of room for improvements in computational efficiency.

```
function [PathList] = enumeratePaths(X)
% Usage enumeratePaths(X)
%
% - X is the sociomatrix (adjacency matrix) of a network (graph);
%   This procedure enumerates all paths of length 1 (essentially
%   the edge list) to length (n-1) if such paths exist.
% - X may be a symmetric (asymmetric) graph (digraph).
% - The output (PathList) is a cell array where PathList{n} contains
%   the listing of all paths of length n.
%
% - jth / last modified on 23 NOV 05

tic % - begin time stamp
    [nR nC] = size(X);  % get dimensions of sociomatrix
%   mpl = nR - 1;       % maximum path length (mpl) is (n-1)
    mpl = 6;
%   paths of length 1...    (essentially the edge list)
    [f t] = find(X);
    PathList{1} = [f t];
%   find paths of length 2 to (n-1) or (mpl)
    for p_length = 2:mpl
      % initialize path list of length p_length
        PathList{p_length} = [];
      % initialize number in current path list
        PathCount = 1;
```

```
      % temporary array containing path list of previous distance
        tmpPrevList = PathList{(p_length-1)};
      % number of paths and columns in path list of previous distance
        [nPrevList colPrevList] = size(tmpPrevList);
      % for each path in previous list, see if it can continue...
        for i = 1:nPrevList
          % get the actor index of the last in line for current path
            nextStart = tmpPrevList(i,colPrevList);
          % find the possible next steps that can be made
            potentialEnd = find(X(nextStart,:));
          % index of how many possible next steps exist
            npe = nnz(potentialEnd);
          % run through possibilities;
          % if an edge can be added, do so...
            for j = 1:npe
              % if this actor has not already been visited
              % in the current path, add him
                if ismember(potentialEnd(j), tmpPrevList(i,:))==0 ;
                    PathList{p_length}(PathCount,:) =
                         [tmpPrevList(i,:) potentialEnd(j)];
                    PathCount = PathCount + 1;
                end % (if) check for already visited node and update
            end % (j) check for possible edges to add (nodes to visit)
        end % (i) action on i-th path in previous path-length list
      % if no more moves/updates are possible, end the program
%         if nnz(PathList{p_length}) == 0
%            break
%         end
        p_length
    end % (p_length)
% - end time stamp (tells user how long it took this program to run)
toc
```

Appendix K. Code: netflowCent.m

This MATLAB function calculates the network flow centrality measure developed by Freeman et al. [1991], given a sociomatrix and the capacity of influence that may flow between each possible (i, j) pair of individuals. Although the sample, taken from [Freeman et al., 1991], indicates a symmetric network, both in connections and arc capacities, this code can accept asymmetric inputs for both categories of data.

```
function [ sna_flowcent ] = netflowCent( X, U )
%   Useage netflowCent(X, U)
%   X = Adjacency matrix (symmetric or asymmetric)
%   U = Matrix capturing upper bounds -- (i,j) entries must match X
%   Assumes lower bound is zero for all arcs
%
% Data from netflow article (Freeman, et al)
% X = [0 1 1 1 0
%      1 0 1 0 0
%      1 1 0 1 1
%      1 0 1 0 0
%      0 0 1 0 0];
%
% U = [0 3 1 2 0
%      3 0 3 0 0
%      1 3 0 2 2
%      2 0 2 0 0
%      0 0 2 0 0];
% Answer with X and U above is...
%    1.0000    7.0000   20.0000    0.3500
%    2.0000    5.0000   20.0000    0.2500
%    3.0000   13.0000   20.0000    0.6500
%    4.0000    6.0000   24.0000    0.2500
%    5.0000    0.0000   30.0000    0.0000
tic
% n = number of nodes (and number of rows) for A
  [n m] = size(X);
% Develop A matrix from X...
  xij = nnz(X);             % xij = number of edges (columns) for A
  A = zeros(n,xij);         % initialize
  [pos neg] = find(X > 0);  % index for +/-1
```

```
  for i = 1:xij             % update A
       A(pos(i),i) = 1;
       A(neg(i),i) = -1;
  end
%   develop s-t pairs to solve for...
    nstPairs = nchoosek(n,2);      % number of solution pairs required
    stPairs = nchoosek((1:n),2);   % index of solution pairs
%   objective function (note -1 since linprog always minimizes)
    f = [zeros(1,xij) -1]';
%   lower bounds (always assumed to be zero
    lb = zeros(1,xij)';
%   upper bounds
    if isempty(U)
        ub = ones(1,xij)';
    else
        uidx = find(U>0);
        ub = U(uidx)';
    end % if isempty...
%   right-hand side
    beq = [zeros(1,n)]';
%   initialize place to store results...
    results = [];
    tstData = [];
    objfns = [];
%   for each possible s-t pair t>s, solve
    for stflow = 1:nstPairs
        s = stPairs(stflow,1);
        t = stPairs(stflow,2);
        fCol = zeros(n,1);
        fCol(s) = -1;
        fCol(t) = 1;
        Aeq = [A fCol];
        [x, fval] = linprog(f,[],[],Aeq,beq,lb,ub);
        objfns = [objfns; [s t x']];
        results = [results; [s t 0 -fval 0]];
%       for each possible s-t pair t>s solution,
%       remove all other nodes other than s or t to ascertain the flow
        for killNode = 1:n  % for all nodes...
%           if it is neither s nor t
            if killNode ~= s & killNode ~= t
%               find cols associated with this node
                killCols = find(Aeq(killNode,:)~=0);
```

```
                tmpAeq = Aeq;
%               remove those columns
                tmpAeq(:,killCols)=[];
%               remove that row
                tmpAeq(killNode,:) = [];
%               adjust the length of the objective function, f
                tmpf = [zeros(1,(xij-nnz(killCols))) -1]';
%               adjust the length of the rhs, beq
                tmpbeq = [zeros(1,(n-1))]';
%               adjust the length of the lower bounds, lb
                tmplb = zeros(1,(xij-nnz(killCols)))';
%               adjust the length of the upper bounds, ub
                tmpub = ub;
                tmpub(killCols) = [];
%               resolve on new graph with s-t and killNode removed...
                [kx, kfval] =
                   linprog(tmpf,[],[],tmpAeq,tmpbeq,tmplb,tmpub);
%               store results
                results =
                   [results; [s t killNode -kfval (-fval + kfval)]];
            end %if killNode...
        end % for killNode...
    end % for stflow
%   calculate measure for each actor....
%   get index of number of solution pairs to facilitate measure
    [solPairs, c] = size(results);
    for i = 1:n
        tmpDenom = 0;
        tmpNumer = 0;
        for j = 1:solPairs
%           denominator data
            if results(j,1)~=i & results(j,2)~=i & results(j,3)==0
                tmpDenom = tmpDenom + results(j,4);
            end % if results(j, 1)...
%           numerator data
            if results(j,3) == i
                tmpNumer = tmpNumer + results(j,5);
            end % if results(j,3)...
        end % for j = ...
%       Layout for tstData is ...
%       Col 1: Actor ID (i)
%       Col 2: Flow that must pass through i from...
```

```
%       Col 3: Sum of all s-t max flows where i is neither s nor t
%       Col 4: (Col 2) / (Col 3) the centrality measure for i
        tstData = [tstData; [i tmpNumer tmpDenom (tmpNumer/tmpDenom)]];
    end % for i = ...
    sna_flowcent = tstData;
toc
```

Appendix L. Code: gnfCent.m

This MATLAB function calculates generalized network flow centrality measure developed in chapter VI. This measure is based upon the centrality measure posited by Freeman et al. [1991], but also incorporates the interpersonal gain multipliers as discussed in chapter VII.

```
function [ tstData ] = gnfCent( G, U )
%   Useage gnfCent(G, U)
%   G = Gains (or loss) matrix -- (i,j) entries must match X
%   U = Matrix capturing upper bounds -- (i,j) entries must match X
%   Assumes lower bound is zero for all arcs
    tic
%   n = number of nodes (and number of rows) for A
    [n m] = size(U);
    look = 0;
%   Develop A matrix from X...
% ***** Still need to incorporate G once this is working well *****
  xij = nnz(U);              % xij = number of edges (columns) for A
  A = zeros(n,xij);          % initialize
%  [pos neg] = find(X > 0);  % index for +/-1
  [pos neg] = find(G > 0);  % index for +/-1
  for i = 1:xij              % update A
        A(pos(i),i) = 1;
        A(neg(i),i) = -G(pos(i),neg(i));
  end
%   develop s-t pairs to solve for...
    nstPairs = nchoosek(n,2);       % number of solution pairs required
     stPairs = nchoosek((1:n),2);   % index of solution pairs
%   objective function (note -1 since linprog always minimizes)
    f = [zeros(1,xij) -1]';
%   lower bounds (always assumed to be zero
    lb = zeros(1,xij)';
%   upper bounds
    if isempty(U)
        ub = ones(1,xij)';
    else
        uidx = find(U>0);
        ub = U(uidx)';
    end % if isempty...
```

```
%   right-hand side
    beq = [zeros(1,n)]';
%   initialize place to store results...
    results = [];
    tstData = [];
    objfns = [];
%   for each possible s-t pair t>s, solve
    for stflow = 1:nstPairs
        s = stPairs(stflow,1);
        t = stPairs(stflow,2);
        fCol = zeros(n,1);
        fCol(s) = -1;
        fCol(t) = 1;
        Aeq = [A fCol];
        [x, fval] = linprog(f,[],[],Aeq,beq,lb,ub);
        objfns = [objfns; [s t x']];
        results = [results; [s t 0 -fval 0]];
%       for each possible s-t pair t>s solution,
%       remove all other nodes
%       other than s or t to ascertain the flow
        for killNode = 1:n  % for all nodes...
%           if it is neither s nor t
            if killNode ~= s & killNode ~= t
%               find cols associated with this node
                killCols = find(Aeq(killNode,:)~=0);
                tmpAeq = Aeq;
%               remove those columns
                tmpAeq(:,killCols)=[];
%               remove that row
                tmpAeq(killNode,:) = [];
%               adjust the length of the objective function, f
                tmpf = [zeros(1,(xij-nnz(killCols))) -1]';
%               adjust the length of the rhs, beq
                tmpbeq = [zeros(1,(n-1))]';
%               adjust the length of the lower bounds, lb
                tmplb = zeros(1,(xij-nnz(killCols)))';
%               adjust the length of the upper bounds, ub
                tmpub = ub;
                tmpub(killCols) = [];
%               resolve on new graph with s-t and killNode removed...
                [kx, kfval] =
                     linprog(tmpf,[],[],tmpAeq,tmpbeq,tmplb,tmpub);
```

```
%               store results
                results =
                     [results; [s t killNode -kfval (-fval + kfval)]];
                look = look + 1
            end %if killNode...
        end % for killNode...
    end % for stflow
%   calculate measure for each actor....
%   get index of number of solution pairs to facilitate measure
    [solPairs, c] = size(results);
    for i = 1:n
        tmpDenom = 0;
        tmpNumer = 0;
        for j = 1:solPairs
%           denominator data
            if results(j,1)~=i & results(j,2)~=i & results(j,3)==0
                tmpDenom = tmpDenom + results(j,4);
            end % if results(j, 1)...
%           numerator data
            if results(j,3) == i
                tmpNumer = tmpNumer + results(j,5);
            end % if results(j,3)...
        end % for j = ...
%       Layout for tstData is ...
%       Col 1: Actor ID (i)
%       Col 2: Flow that must pass through i from...
%       Col 3: The sum of all s-t max flows where i is neither s nor t
%       Col 4: (Col 2) / (Col 3)... the centrality measure for i
        tstData = [tstData; [i tmpNumer tmpDenom (tmpNumer/tmpDenom)]];
    end % for i = ...
toc
```

Appendix M. KPP-2 and Layered Networks

The underlying assumption applicable to the KPP-2 methodologies developed is that the network of interest is connected. When applying the key player methodologies developed in this dissertation to layered networks, not all layers are guaranteed to be connected. In addition, not all layers may share the same set of actors. Therefore, slight modifications to the constraint matrix must be considered. The code presented in this appendix facilitates the analysis methodology described in Section 5.7.

The following function, buildKPPA.m, takes a sociomatrix of a given layer as input, and provides the requisite constraint matrix for the KPP-2 applications presented within this research.

```
function [kppA] = buildKPPA(X)
%
[N,m] = size(X);
notavail = find(sum(X)==0);
kppA = -(X + eye(N));
kppA(notavail,:) = [];
```

Assuming a symmetric sociomatrix, this function prevents the inclusion of rows associated with actors that do not have interpersonal relationships within the given context of interest. If the sociomatrix is asymmetric, then $\mathbf{X}'$ must be input into this function.

To account for multiple layers simultaneously, the output of this function for each layer must be horizontally concatenated, forming the multi-layer constraint matrix for input into a modified version of the KPP-2 programs presented. Note that this approach currently presented assumes a reach of one-step. If a reach greater than one between a key player and its assigned actors is desired, the analyst should input a dichotomized version of the 2-step reachability matrix, which can be determined by the code presented in Appendix C. An example of the modification required for the NRm.m code is provided below.

```
function [ kp_nr ] = NR_layers( A , enum )
% Useage NR_layers(A, enum)
%     A - Multi-layer constraint matrix
%  enum - Enumerate solutions until infeasible or maximum
%         number have been found
%         (maxsol is currently set at 100 solutions)
%  Requires the MATLAB Optimization Toolbox
%  This determines the minimum number of key players required to
%  cover multiple relationship layers.  The input A must be
%  developed prior to implementing this function.
    tic
%   ***** initialize variables *****
    maxSol = 100;       % maximum number of solutions (for enum = 1)
    numSol = 0;         % current number of solutions
    [N, M] = size(A);      % dimension of X
    b = - ones(N,1);    % column vector as RHS
    f = ones(M,1);      % column vector as obj. function coeff.
    kp_nr = [];         % storage for solutions
    firstSol = 1;       % boolean flag to control enumeration loops
    foundall = 0;       % boolean flag to control enumeration loops
    kpK = 0;            % kp-set size, facilitates enumeration loops
    if enum
        while not(foundall)
%           column vector of x as solution
            [mpSol, obj, flag] = bintprog(f, A, b);
            if firstSol
                kpK = obj;      % size of min. dom. set
                firstSol = 0;
            end % if firstSol...
            if flag == -2 | (numSol == maxSol) | (obj ~= kpK)
                foundall = 1
            else
                kp_set = find(mpSol==1);
                kp_nr = [kp_nr kp_set];
%               add extra rhs
                b = [b; (kpK - 1)];
%               add extra constraint
                A = [A; mpSol'];
                numSol = numSol + 1;
            end % if flag...
        end % while not...
%       Plot data in histogram (only when enumerating solutions)
```

```
        [a, b] = size(kp_nr);
        pdata = [];
        xaxis = 1:N;
        for j = 1:a
            pdata = [pdata kp_nr(j,:)];
        end
        hist(pdata,xaxis);
        xlim([0 N]);
        ylim([0 b]);
    else
%       column vector of x as solution
        mpSol = bintprog(f, A, b);
        kp_nr = find(mpSol==1) % key players
    end % if enum...
toc     % time elapsed
```

Appendix N. Pair-wise Gains Process

This appendix details the process used to develop pair-wise gain multipliers. The steps of the process include:

1. Collect personal attribute data on individuals of interest. This data must comprise components of the charisma model depicted in Figure 7.2.
2. Determine which individuals are charismatic (combination of authority and persuasive influence) and which are not. Assign a response variable 1 (0) for charismatic (non charismatic) individuals.
3. Use logistic regression to determine a model sufficient for use in obtaining probability estimates $P(Y_i = 1|\mathbf{X})$.
4. Use vector of probability estimates and appropriate sociomatrix as input to the MATLAB function provided below.
5. Output of `generateG` is an $n \times n$ matrix of gain multipliers g_{ij} that can be used as input for the generalized network flow centrality measure provided in Appendix L or as input to the influence course of action analysis process described in Section 7.3

```
function [G] = generateG(ep, X)
% Useage, generateG(ep, X) generates a gain matrix for use in the
% gnfCent( G, U ) generalized network flow centrality measure.
% ep is a vector of estimated probabilities of a set of actors
% based upon a logistic regression
% G is the gain multiplier matrix as applied to all pair-wise
% links in X
[n, m] = size(X); G = zeros(n,n); noX = find(X==0); [i, j] =
find(G==0); idx = [i j]; for k=1:(size(idx,1))
   G(idx(k,1),idx(k,2)) = 1+ep(idx(k,1))-ep(idx(k,2));
end G(noX) = 0;
```

Appendix O. JI Member Data

This appendix provides the descriptive data of the 48 JI members of interest used in various portions of this research. The network and attribute data are derived from Sageman's book, Understanding Terror Networks, published in 2004. Tables 15.1 and 15.2 provide the index and list of study names, full names, age of the member when he joined JI, and the year when the member joined JI. The latter two columns present the data used in the logistic regression approach to measuring interpersonal gains (or losses) of influence.

Table 15.1: JI Membership (Subset of 48 Actors)

Index	Study Name	Full Name	Year Joined	Age Joined
1	Baasyir	Abu Bakar Baasyir	1989	51
2	Sungkar	Abdullah Sungkar	1989	52
3	Hambali	Encep Nurjaman	1989	25
4	Mukhlas	Ali Ghufron bin Nurhasyim	1989	29
5	Iqbal	Fikiruddin Muqti	1989	30
6	Faruq	Omar al-Faruq	1991	20
7	Syawal	Yassin Syawal	1989	24
8	Ghozi	Fathur Rahman al-Ghozi	1989	18
9	Samudra	Abdul Aziz	1991	21
10	Jabir	Enjang Bastaman	1991	31
11	Amrozi	Amrozi bin Nurhasyim	1992	30
12	Imron	Ali Imron bin Nurhasyim	1990	18
13	Sufaat	Yazid Sufaat	1998	33
14	Dwikarna	Agus Dwikarna	1990	26
15	Mobarok	Hutomo Pamungkus	1990	20
16	Yunos	Saifullah Yunos	1989	19
17	Mistooki	Jafaar bin Mistooki	1990	29
18	Faiz	Faiz bin Abu Bakar Bafana	1991	29
19	Hasyim	Hasyim bin Abbas	1991	30
20	Sulaeman	Mohammed Nasir bin Abbas	1989	20
21	Hussein	Abdul Rahman Ayub	1989	23
22	Ayub	Abdul Rahim Ayub	1989	23
23	Azahari	Azahari bin Husin	1990	33
24	Zulkarnaen	Aris Sumarsomo	1989	26

Table 15.2: JI Membership (Subset of 48 Actors)

Index	Study Name	Full Name	Year Joined	Age Joined
25	Ghoni	Suranto Abdul Ghoni	1991	23
26	Top	Noordin Mohammad Top	1990	20
27	Idris	Jhoni Hendrawan	1993	23
28	Mustofa	Pranata Yudha	1989	27
29	WanMin	Wan Min bin Wan Mat	1990	29
30	Maidin	Ibrahim bin Maidin	1989	39
31	Sani	Asmar Latin Sani	1993	18
32	Dulmatin	Umar Dul Matin	1990	20
33	Farik	Mohammad Farik bin Amin	1993	26
34	Lillie	Bashir bin Lap	1993	26
35	Yunos2	Muhammad Amin Mohamed Yunos	1999	17
36	Naharudin	Muhammad Arif Naharudin	2000	17
37	Gungun	Rusman Gunawan	1999	23
38	Marzuki	Zulkifli Marzuki	1990	26
39	Kastari	Mas Selamat bin Kastari	1990	29
40	Hafidh	Mohammed Faiq bin Hafidh	1989	31
41	Setiono	Edy Setiono	1989	28
42	BinHir	Zulkifli bin Hir	1991	25
43	Rusdan	Thoriqudin	1989	29
44	Mustaqim	Mustaqim	1989	28
45	Fathi	Fathi Abu Bakar Bafana	1991	26
46	Khalim	Mohamed Khalim bin Jaffar	1993	31
47	Roche	Jack Roche	1998	45
48	Thomas	Jack Thomas	2001	27

Bibliography

Z. Abuza. Abu dujana: Jemaah Islamiyah's new al-Qaeda linked leader. *Terrorism Focus*, 3(13):4–5, April 2006.

R. Ahuja, T. Magnangti, and J. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice Hall, Upper Saddle River, 1993.

R. Albert, H. Jeong, and A. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

C. P. Alderfer. An empirical test of a new theory of human needs. *Organizational Behavior and Human Performance*, 4:142–175, 1969.

F. Amblard and G. Deffuant. The role of network topology on extremism propagation with the relative agreement opinion dynamics. *Physica A*, 343:725–738, 2004.

C. Anderson, O. P. John, D. Keltner, and A. M. Kring. Who attains social status? effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology*, 81(1):116–132, 2001.

Anonymous. Fundamental analysis, 27 Jan 2005.

T. M. Apostol. *Mathematical Analysis*. Addison Wesley Longman, Reading, 1974.

R. Axtell. Why agents? on the varied motivations for agent computing in the social sciences. Working Paper Working Paper No. 17, 2000.

W. E. Baker and R. R. Faulkner. The social organization of conspiracy: Illegal networks in the heavy electrical equipment industry. *American Sociological Review*, 58(6):837–860, 1993.

A. Barabási. *Linked: The New Science of Networks*. Perseus Publishing, Cambridge, 2002.

A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, Oct 1999.

A. Barabási and E. Bonabeau. Scale-free networks. *Scientific American*, pages 50–59, 2003.

F. H. Barron and B. E. Barrett. Decision quality using ranked attribute weights. *Management Science*, 42(11):1515–1523, 1996.

J. Baumes, M. Goldberg, M. Magdon-Ismail, and W. Al Wallace. Discovering hidden groups in communication networks. *Lecture Notes in Computer Science*, 3073:378–389, Jan 2004.

M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali. *Linear Programming and Network Flows*. John Wiley & Sons, New York, 1990.

L. Berry, G. E. Curtis, R. A. Hudson, and N. A. Kollars. A global overview of narcotics-funded terrorist and other extremist groups. Technical report, Federal Research Division, Library of Congress, 2002.

E. S. Bogardus. Measuring social distances. *Journal of Applied Sociology*, 9:299–308, 1925.

J. M. Bolland. Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, 10:233–253, 1988.

P. Bonacich, A. C. Holdren, and M. Johnston. Hyper-edges and multidimensional centrality. *Social Networks*, 26:189–203, 2004.

P. Bonacich and P. Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23:191–201, 2001.

P. B. Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, 92:1170–1182, 1987.

S. P. Borgatti. *Identifying sets of key players in a network. International Conference Integration of Knowledge Intensive Multi-Agent Systems*. National Research Council, 2003a.

S. P. Borgatti. Key player. Analytic Technologies: Boston, 2003b.

S. P. Borgatti. Centrality and network flow. *Social Networks*, 27:55–71, 2005.

S. P. Borgatti. Identifying sets of key players in a social network. *Computational and Mathematical Organization Theory*, 12(1):21–34, 2006.

S. P. Borgatti, K. M. Carley, and D. Krackhardt. On the robustness of centrality measures under conditions of imperfect data. *Social Networks*, 28(2):124–136, 2006.

S. P. Borgatti and M. G. Everett. A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.

S. P. Borgatti, M. G. Everett, and L. C. Freeman. *UCINET for Windows: Software for Social Network Analysis.* Analytic Technologies, Harvard, MA, 2002.

P. A. Bottomley and J. R. Doyle. A comparison of three weight elicitation methods: good, better, and best. *Omega*, 29:553–560, 2001.

B. Bozkaya, J. Zhang, and E. Erkut. *An efficient genetic algorithm for the p-median problem*, chapter Chapter 6, pages 179–205. Springer-Verlag, 2002.

U. Brandes and D. Fleischer. Centrality measures based on current flow. In *STACS 2005: 22nd Annual Symposium on Theoretical Aspects of Computer Science*, volume 3404, pages 533–544, Stuttgart, 24Feb2005 2005. Springer-Verlag GmbH.

D. J. Brass. A social network perspective on human resources management. *Research in Personnel and Human Resources Management*, 13:39–79, 1995.

D. D. Brewer and C. M. Webster. Forgetting of friends and its effects on measuring friendship networks. *Social Networks*, 21:361–373, 1999.

M. Buchanan. *Nexus: Small Worlds and the Groundbreaking Theory of Networks.* W. W. Norton, New York, 2002.

R. S. Burt and T. Schøtt. Relation contents in multiple networks. *Social Science Research*, 14:287–308, 1985.

J. R. Busemeyer. *Dynamic Decision Making*, volume 6 of *International Encyclopedia of the Social and Behavioral Sciences*, pages 3903–3908. Elsevier Press, Oxford, 2002.

K. M. Carley. Dynamic network analysis. In R. Breiger, K. Carley, and P. Pattison, editors, *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 133–145, Washington D. C., 2003. The National Academies Press.

K. M. Carley. Organizational risk analyzer. Software, May 2006.

K. M. Carley, J. Lee, and D. Krackhardt. Destabilizing networks. *Connections*, 24 (3):79–92, 2002.

T. Carpenter, G. Karakostas, and D. Shallcross. Practical issues and algorithms for analyzing terrorist networks, 2002.

C. Carroll. Canonical correlation analysis: Assessing links between multiplex networks. *Social Networks*, 28:310–330, 2006.

N. Christofides. *Graph Theory: An Algorithmic Approach*. Academic Press, New York, 1975.

C. Clark. Modeling and analysis of clandestine networks. Master's thesis, Air Force Institute of Technology, 2005.

R. T. Clemen. *Making hard decisions: An introduction to decision analysis*. Duxbury Press, Pacific Grove, 1996.

E. Costenbader and T. Valente. The stability of centrality measures when networks are sampled. *Social Networks*, 25:283–307, 2003.

D. L. Costley, C. Santana-Melgoza, and R. Todd. *Human Relations in Organizations*. West Publishing Company, Minneapolis, 1994.

N. D. Curet, J. DeVinney, and M. E. Gaston. An efficient network flow code for finding all minimum cost s-t cutsets. *Computers & Operations Research*, 29:205–219, 2002.

Cyram. *NetMiner II*, volume Ver. 2.5.0. Cyram Co., Ltd, Seoul, 2004.

A. Degenne and M. Forsé. *Introducing Social Networks*. Introducing Statistical Methods. SAGE Publications Ltd., London, 1st edition, 1999.

I. DeGraw. A dynamic valuation model. *On Wall Street*, 11(5):78, 2001.

N. Deo. *Graph Theory with Applications to Engineering and Computer Science*. Prentice-Hall, Inc., Englewood Cliffs, 1974.

DHS. National strategy for homeland security. Technical report, Department of Homeland Security, 2002.

DOD. Joint Publication 3-0: Doctrine for Joint Operations, 2001.

DOD. Joint Publication 3-60: Joint Doctrine for Targeting, Jan 2002.

DOD. Joint Publication 3-53: Doctrine for Joint Psychological Operations, Sep 2003.

DOD. Joint Publication 1-02: Department of Defense Dictionary of Military and associated terms, 2005.

DOD. Joint Publication 3-13: Information Operations, Feb 2006a.

DOD. Quadrennial Defense Review Report, Feb 2006b.

M. J. Dombroski and K. M. Carley. NETEST: Estimating a terrorist network's structure: Graduate student best paper award, CASOS 2002 conference. *Computational & Mathematical Organization Theory*, 8:235–241, 2002.

P. Domingos and M. Richardson. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, New York, NY, USA, 2001. ACM Press.

P. Doreian and K. L. Woodward. Fixed list versus snowball selection of social networks. *Social Science Research*, 21:216–233, 1992.

D. D. Downs. Gauging the commitment of clandestine group members. Master's thesis, Air Force Institute of Technology, 2006.

S. Driscoll. De-mytstifying extremism, 2005.

J. S. Dyer and R. K. Sarin. Measurable multiattribute value functions. *Operations Research*, 27:810–822, 1979.

W. Edwards. Dynamic decision-theory and probabilistic information-processing. *Human Factors*, 4(2):59–73, 1962.

W. Edwards. How to use multiattribute utility measurement for social decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(5):326–339, 1977.

W. Edwards and F. H. Barron. Smarts and smarter: Improved simple methods for multiattribute utility measurement. *Organizational Behavior and Human Decision Processes*, 60(3):306–325, 1994.

H. C. Ellis and R. R. Hunt. *Fundamentals of Cognitive Psychology.* Brown & Benchmark, Madison, 1993.

D. Eppstein and J. Wang. Fast approximation of centrality. *Journal of Graph Algorithms and Applications*, 8(1):39–45, 2004.

P. Erdos and A. Renyi. On random graphs. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

M. G. Everett and S. P. Borgatti. The centrality of groups and classes. *Journal of Mathematical Sociology*, 23(3):181–201, 1999.

P. V. Fellman and R. Wright. Modeling terrorist networks - complex systems at the mid-range. In *Complexity, ethics and creativity conference*, 18 Sep 2003.

K. A. Frank and J. Y. Yasumoto. Linking action to social structure within a system: Social capital within and between subgroups. *American Journal of Sociology*, 104 (3):642–686, 1988.

L. C. Freeman. The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity*, 14:585–592, 1980.

L. C. Freeman, S. P. Borgatti, and D. R. White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks*, 13(2):141–154, 1991.

J. R. French. A formal theory of social power. *Psychological Review*, 63:181–184, 1956.

N. E. Friedkin. A test of structural features of granovetter's strength of weak ties theory. *Social Networks*, 2:411–422, 1980.

N. E. Friedkin. A formal theory of social power. *Journal of Mathematical Sociology*, 12(2):103–126, 1986.

N. E. Friedkin. A guttman scale for the strength of an interpersonal tie. *Social Networks*, 12:239–252, 1990.

N. E. Friedkin and K. S. Cook. Peer group influence. *Sociological Methods & Research*, 19(1):122–143, 1990.

N. E. Friedkin and E. C. Johnsen. Control loss and fayol's gangplanks. *Social Networks*, 24:395–406, 2002.

W. F. Gabrielli and D. von Winterfeldt. Are importance weights sensitive to the range of alternatives in multiattribute utility measurement. Technical Report 78-6, 1978.

T. Gal. *Potoptimal Analyses, Parametric Programming, and Related Topics*. McGraw-Hill, Inc., Great Britain, 1979.

Rajiv Gandhi, Samir Khuller, and Aravind Srinivasan. Approximation algorithms for partial covering problems. *Journal of Algorithms*, 53:55–84, 2004.

Herbert Goldhamer and Edward A. Shils. Types of power and status. *The American Journal of Sociology*, 45(2):171–182, 1939.

L. A. Goodman. Snowball sampling. *Annals of Mathematical Statistics*, 32:148–170, 1961.

R. V. Gould. Multiple networks and mobilization in the paris commune, 1871. *American Sociological Review*, 56(6):716–729, 1991.

F. Grandoni. A note on the complexity of minimum dominating set. *Journal of Discrete Algorithms*, 4:209–214, 2006.

M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78 (6):1360–1380, 1973.

M. S. Granovetter. Network sampling: Some first steps. *American Journal of Sociology*, 81(6):1267–1303, 1976.

M. S. Granovetter. Threshold models of collective behavior. *The American Journal of Sociology*, 83(6):1420–1443, May 1978.

M. S. Granovetter. *Sociological Theory*, volume 1, chapter 7, The strength of weak ties: A network theory revisited, pages 201–233. 1983.

M. S. Granovetter and R. Soong. Threshold models of diffusion and collective behavior. *Journal of Mathematical Sociology*, 9:165–179, 1983.

H. J. Greenberg. *Optimal attack of a command and control communications network*. PhD thesis, The Johns Hopkins University, 1968.

R. P. Hämäläinen and A. A. Salo. The issue is understanding the weights. *Journal of Mutli-Criteria Decision Analysis*, 6:340–343, 1997.

G. Y. Handler and P. B. Mirchandani. *Location on Networks: Theory and Algorithms*. MIT Press, Cambridge, MA, 1979.

L. Harris. Al Qaeda's fantasy ideology. *Policy Review*, 114:19–37, 2002.

R. V. Hartley. *Operations Research: A Managerial Emphasis*. Goodyear Publishing Company, Inc., 1976.

C. Haythornthwaite. A social network theory of tie strength and media use: A framework for evaluating multi-level impacts of new media. Technical Report UIUCLIS–2002/1+DKRC, University of Illinois at Urbana-Champaign, 1999.

F. Herzberg, B. Mausner, and B. B. Snyderman. *The motivation to work*. Wiley, New York, 1965.

F. S. Hillier and G. J. Lieberman. *Introduction to Operations Research*. McGraw-Hill, Inc., 6th edition, 1995.

J. M. Hite. Patterns of multidimensionality among embedded network ties: a typology of relational embeddedness in emerging entrepreneurial firms. *Strategic Organization*, 1(1):9–49, 2003.

C. H. Hubbell. An input-output approach to clique identification. *Sociometry*, 28: 377–399, 1965.

R. A. Hudson. The sociology and psychology of terrorism: Who becomes a terrorist and why? Technical report, Library of Congress, 1999.

J. P. Ignizio. *Linear Programming in Single- & Multiple-Objective Systems*. Prentice-Hall, Inc., Englewood Cliffs, 1982.

S. L. Jack. The role, use and activation of strong and weak network ties: A qualitative analysis. *Journal of Management Studies*, 42(6):1233–1259, 2005.

J. J. Jarvis. *Optimal attack and defense of a command and control communications neetwork*. PhD thesis, The Johns Hopkins University, 1968.

R. S. John, W. Edwards, and L. Collins. A comparison of importance weights for multiattribute utility analysis derived from holistic, indifference, direct subjective and rank order judgments. Technical Report 80-4, 1980.

M. Johns and B. G. Silverman. How emotions and personality effect the utility of alternative decision: A terrorist target selection case study. In *10th Conference on Computer Generated Forces & Behavioral Representations*, 2001.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18 (1):39–43, 1953.

R. L. Keeney. Building models of values. *European Journal of Operational Research*, 37:149–157, 1988.

R. L. Keeney. Common mistakes in making value trade-offs. *Operations Research*, 50(6):935–945, 2002.

R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. Cambridge University Press, Cambridge, 1993.

R. L. Keeney and D. von Winterfeldt. On the uses of expert judgement on complex technical problems. *IEEE Transactions on Engineering Management*, 36(2):83–86, 1989.

H. C. Kelman. Processes of opinion change. *Public Opinion Quarterly*, 25:57–78, 1961.

D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM Press.

K. T. Kennedy. An analysis of multi-layered networks. Master's thesis, Air Force Institute of Technology, 2003.

C. W. Kirkwood. *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Duxbury Press, Belmont, 1997.

V. E. Krebs. Mapping networks of terrorist cells. *Connections*, 24(3):43–52, 2002.

D. L. Kreher and D. R. Stinson. *Combinatorial Algorithms: Generation, Enumeration, and Search*. CRC Press, Boca Raton, 1999.

I. H. LaValle. *Fundamentals of Decision Analysis*. Holt, Rinehart, and Winston, New York, 1978.

R. Th. A. J. Leenders. Modeling social influence through network autocorrelation: Constructing the weight matrix. *Social Networks*, 24:21–47, 2002.

J. A. Leinart. A network disruption modeling tool. Master's thesis, Air Force Institute of Technology, 1998.

J. A. Leinart, R. F. Deckro, J. M. Kloeber, and J. A. Jackson. A network disruption modeling tool. *Military Operations Research*, 7(1):69–77, 2002.

J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the 11th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, pages 177–187, New York, NY, 2005. ACM Press.

D. Z. Levin, R. Cross, and L. C. Abrams. The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. In *Academy of Management Proceedings*, 2002.

H. Li, L. Li, J. Wang, Z. Mo, and Y. Li. Fuzzy decision making based on variable weights. *Mathematical and Computer Modeling*, 39:163–179, 2004.

D. Liben-Nowell. *An Algorithmic Approach to Social Networks*. PhD thesis, Massachusetts Institute of Technology, 2005.

D. A. Lind, W. G. Marchal, and R. D. Mason. *Statistical Techniques in Business & Economics*. McGraw-Hill Irwin, Boston, 2002.

R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data*. John Wiley & Sons, 1987.

F. A. Lootsma. *Multi-criteria decision analysis via ratio and difference judgement*. Kluwer Academic Publishers, Boston, 1999.

L. Lopez, J. Mendes, and M. Sanjuan. Hierarchical social networks and information flow. *Physica A*, 316:695–708, 2002.

M. J. Lovaglia, R. Willer, and L. Troyer. Power, status, and collective action: Developing fundamental theories to address a substantive problem. *Advances in Group Processes*, 20:105–131, 2003.

J. Ma, Z. Fan, and Q. Wei. Existence and construction of weight-set for satisfying preference orders of alternatives based on additive multi-attribute value model. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 31(1):66–72, 2001.

E. Mansfield. *Microeconomics*. W. W. Norton & Company, Inc., New York, 1994.

P. V. Marsden and K. E. Campbell. Measuring tie strength. *Social Forces*, 63: 482–501, 1984.

M. Marsili, F. Vega-Redondo, and F. Slanina. The rise and fall of a networked society: A formal model. *PNAS*, 101(6):1439–1442, 2004.

R. K. Martin. *Large Scale Linear and Integer Optimization: A Unified Approach*. Kluwer Academic Publishers, 1999.

R. C. Mayer, J. H. Davis, and F. D. Schoorman. An integration model of organizational trust. *The Academy of Management Review*, 20:709–734, 1995.

N. Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming*, 7:97–107, 1974.

Y. W. Meng. Bali in the shadow of terror. *Asia Times Online*, 2004.

M. Migliore, V. Martorana, and F. Sciortino. An algorithm to find all paths between two nodes in a graph. *Journal of Computational Physics*, 87(1):231–236, 1990.

S. Milgram. The small world problem. *Psychology Today*, 22:61–67, 1967.

G. L. Miller and J. S. Naor. Flow in planar graphs with multiple sources and sinks. *SIAM Journal on Computing*, 24(5):1002–1017, 1995.

Ministry of Home Affairs, Singapore. Singapore government press statement on ISA arrests, Jan 2002. URL `http://www2.mha.gov.sg/mha/`.

R. B. Misra and K. B. Misra. Enumeration of all simple paths in a communication network. *Microelectronics and Reliability*, 20(4):419–426, 1980.

P. R. Monge and N. S. Contractor. *Theories of Communication Networks*. Oxford University Press, Oxford, 2003.

J. L. Moreno. *Who shall survive? A new approach to the problem of human interrelations.* Beacon House, Inc., Beacon, 1953.

S. Mussi. Putting value of information theory into practice: a methodology for building sequential decision support systems. *Expert Systems*, 21(2):92–103, 2004.

L. Narens and R. D. Luce. Measurement: The theory of numerical assignments. *Psychological Bulletin*, 99(2):166–180, 1986.

National Commission on Terrorist Attacks Upon the United States. 9/11 Commision Report. Final report, U. S. Government, 2004. URL `http://www.9-11commission.gov/report/index.htm`.

G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. John Wiley & Sons, Inc., New York, 1999.

J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman. *Applied Linear Statistical Models*. WCB McGraw-Hill, 4 edition, 1996.

M. Newman, D. Watts, and S. Strogatz. Random graph models of social networks. *PNAS*, 99(90001):2566–2572, 2002.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

B. Pabjan. Measuring the social relations: Social distance in social structure–a study of prison community. *Acta Physica Polonica B*, 36(8):2559–2574, 2005.

V. N. Parrillo and C. Donoghue. Updating the Borgardus social distance studies: A new national survey. *The Social Sciences Journal*, 42:257–271, 2005.

K. R. Parthasarathy. Enumeration of paths in digraphs. *Psychometricka*, 29(2): 153–165, 1964.

Lucia D. Penso and Valmir C. Barbosa. A distributed algorithm to find k-dominating sets. *Discrete Applied Mathematics*, 141:243–253, 2004.

R. M. Perloff. *The Dynamics of Persuasion: Communication and Attitudes in the 21st Century*. Lawrence Erlbaum Associates, 2003.

M. S. Pinkstaff. An approach to disrupting communication networks. Master's thesis, Air Force Institute of Technology, 2001.

Gillo Pontecorvo. Battle of algiers, 1967.

J. M. Post. *Military Studies in the Jihad Against the Tyrants: The Al-Qaeda Training Manual*. USAF Counterproliferation Center, Maxwell Air Force Base, 2005.

M. Pöyhönen and R. P. Hämäläinen. On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129:569–585, 2001.

K. A. Pruitt. Modeling homeland security: A value focused thinking approach. Master's thesis, Air Force Institute of Technology, 2003.

J. Reese. Methods for solving the p-median problem: An annotated bibliography. Technical Report 96, Trinity University, 2005.

R. S. Renfro. *Modeling and Analysis of Social Networks*. PhD thesis, Air Force Institute of Technology, 2001.

R. S. Renfro and R. F. Deckro. Network flow approaches to social networks. Working Paper.

R. S. Renfro and R. F. Deckro. A flow model social network analysis of the Iranian government. *Military Operations Research*, 8(1):5–16, 2003.

E. T. Rolls. *Neural Basis of Emotion*, pages 4444–4449. International Encyclopedia of the Social & Behavioral Sciences. Elsevier Science Ltd, Oxford, 2001.

R. Rothenberg. From whole cloth: Making up the terrorist network. *Connections*, 24(3):36–42, 2002.

M. Sageman. *Understanding Terror Networks*. University of Pennsylvania Press, Philadelphia, 2004.

W. S. Sarle. Measurement theory: Frequently asked questions. In *Disseminations of the International Statistical Applications Institute*, volume 1, pages 61–66. ACG Press, Wichita, 4th edition, 1995.

S. Schenkerman. Use and abuse of weights in multiple objective decision support models. *Decision Sciences*, 22(2):369–378, 1991.

C. W. Scherer and H. Cho. A social network contagion theory of risk perception. *Risk Analysis*, 23(2):261–267, 2003.

R. Sedgewick. *Algorithms*. Addison-Wesley Publishing Company, 1984.

S. B. Seidman. Structures induced by collections of subsets: a hypergraph approach. *Mathematical Social Sciences*, 1:381–396, 1981.

J. S. Seiter and R. H. Gass. *Perspectives on Persuasion, Social Influence, and Compliance Gaining*. Pearson, Boston, 2004.

B. Shamir. Social distance and charisma: Theoretical notes and an exploratory study. *Leadership Quarterly*, 6(1):19–47, 1995.

B. G. Silverman, R. Might, R. Dubois, H. Shin, M. Johns, and R. Weaver. Toward a human behavior models anthology for synthetic agent development. In *10th Conference on Computer Generated Forces Proceedings*. SISO & IEEE, 2001.

Stephen Slade. A realistic model of rationality. In Michael Fehling, editor, *Working Papers of the (AAAI) Fall Symposium on Rational Agency: Concepts, Theories, Models, and Applications*, pages 126–130, 1995. URL "citeseer.ist.psu.edu/41894.html".

J. Sounderpandian. Should we teach sensitivity analysis report? *Decision Line*, 32 (5):4–5, 2001.

M. K. Sparrow. The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks*, 13:251–274, 1991.

O. Sporns. *Neuroscience Databases: A Practical Guide*, chapter 12: Graph theory methods for the analysis of neural connectivity patterns, pages 169–183. Klüwer, 2002.

E. Sprinzak. Rational fanatics. *Foreign Policy*, 120:66–73, 2000.

K. Stephenson and M. Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37, 1989.

S. E. Sterling. Aggregation techniques to characterize social networks. Master's thesis, Air Force Institute of Technology, 2004.

T. J. Stewart. Robustness of additive value function methods in mcdm. *Journal of Multi-Criteria Decision Analysis*, 5(4):301–309, 1996.

W. G. Stillwell and W. Edwards. Rank weighting in multiattribute utility decision making: avoiding the pitfalls of equal weights. Research Report 79-2, 1979.

D. Stork and W. D. Richards. Nonrespondents in communication network studies: Problems and possibilities. *Group & Organization Management*, 17(2):193–209, 1992.

The President. The National Security Strategy of the United States of America. Technical report, U. S. Government, 2006.

B. E. Thomason, T. R. Coffman, and S. E. Marcus. Sensitivity of social network analysis metrics to observation noise. In *2004 Aerospace Conference*, volume 5, pages 3206–3215. IEEE, 2004.

M. Tsvetovat and K. M. Carley. Structural knowledge and success of anti-terrorist activity: The downside of structural equivalence. *Journal of Social Structure*, 6 (2):np, 2005.

U. S. Department of State. Country reports on terrorism 2004. Technical report, Office of the Coordinator for Counterterrorism, 2005.

USAF. Air Force Doctrine Document 2-1: Air Warfare, 2000.

T. W. Valente. Social network thresholds in the diffusion of innovations. *Social Networks*, 18:69–89, 1996.

T. W. Valente and R. K. Foreman. Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, 20:89–105, 1998.

K. M. van Meter. Terrorists/liberators: Researching and dealing with adversary social networks. *Connections*, 24(3):66–78, 2002.

R. von Nitzsch and M. Weber. The effect of attribute ranges on weights in multiattribute utility measurements. *Management Science*, 39(8):937–943, 1993.

D. von Winterfeldt and W. Edwards. *Decision analysis and behavioral research*. Cambridge University Press, Cambridge, 1986.

S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

J. K. Watters and P. Biernacki. Targeted sampling: Options for the study of hidden populations. *Social Problems*, 36(4):416–430, 1989.

D. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, 1999.

R. Weaver, B. G. Silverman, H. Shin, and R. Dubois. Modeling and simulating terrorist decision-making: A performance moderator function approach to generating

virtual opponents. In *10th Conference on Computer Generated Forces Proceedings*, New York, 2001. SISO & IEEE.

R. L. Weisbrod, K. B. Davis, and A. Freedy. Adaptive utility assessment in dynamic decision processes: An experimental evaluation of decision aiding. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(5):377–383, 1977.

W. T. Wolters and B. Mareschal. Novel types of sensitivity analysis for additive mcdm methods. *European Journal of Operational Research*, 81:281–290, 1995.

J. J. Xu and C. Hsinchun. Fighting organized crimes: Using shortest-path algorithms to identify associates in criminal networks. *Decision Support Systems*, 38(3):473–487, 2004.

Z. Yamaguchi. The flow of information through social networks: Diagonal-free measures of inefficiency and the structural determinants of inefficiency. *Social Networks*, 16:57–86, 1994.

Song Yang and David Knoke. Optimal connections: Strength and distance in valued graphs. *Social Networks*, 23:285–295, 2001.

Y. Yin and K. Yasuda. Similarity coeffecient methods applied to the cell formation problem: A comparative investigation. *Computers & Industrial Engineering*, 48: 471–489, 2005.

Vita

Major J. Todd Hamill was born in Birmingham, Alabama. Upon graduation from Southside High School in Fort Smith, Arkansas, he enlisted in the United States Air Force as a medical administrative specialist. After just over one year of active duty service, he was accepted to the United States Air Force Academy Preparatory School, and ultimately graduated from the United States Air Force Academy in June of 1993. His first assignment was as an analyst supporting precision guided munitions evaluations. During his second assignment, a B-2 survivability analyst, he completed a Master of Science Degree in Industrial Engineering at New Mexico State University. His third assignment garnered him another Master of Science Degree, in Operations Research, at the Air Force Institute of Technology (AFIT). Prior to entrance into the Doctoral program at AFIT, Major Hamill served as a scientific analyst at the Space and Missile Systems Center at Los Angeles Air Force Base. Major Hamill's follow-on assignment will involve the study of Information Operations and Information Warfare techniques at U.S. Strategic Command at Offutt AFB, Nebraska.

Permanent address: 2950 Hobson Way, Bldg. 640
Wright-Patterson AFB, OH 45433-7765

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
21-12-2006	**Doctoral Dissertation**	Aug 2003 – Sep 2006

4. TITLE AND SUBTITLE	
Analysis of Layered Social Networks	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Hamill, Jonathan Todd, Major, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPAFB OH 45433-7765	AFIT/DS/ENS/06-03

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		**10. SPONSOR/MONITOR'S ACRONYM(S)**
AFRL/HECS Attn: Gregory D. Sullivan, 2Lt, USAF 2689 G Street WPAFB OH 45433-7022 DSN: 785-8015	NASIC/FCEB Attn: Billy D. Darnell, Major, USAF 4180 Watson Way WPAFB OH 45433-5648 DSN: 986-1023	**AFRL/HECS; NASIC/FCEB** **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Prevention of near-term terrorist attacks requires an understanding of current terrorist organizations to include their composition, the actors involved, and how they operate to achieve their objectives. To aid this understanding, operations research, sociological, and behavioral theory relevant to the study of social networks are applied, thereby providing theoretical foundations for new methodologies to analyze non-cooperative organizations, defined as those trying to hide their structure or are unwilling to provide information regarding their operations.

Techniques applying information regarding multiple dimensions of interpersonal relationships, inferring from them the strengths of interpersonal ties, are explored. A layered network construct is offered that provides new analytic opportunities and insights generally unaccounted for in traditional social network analyses. These provide decision makers improved courses of action designed to impute influence upon an adversarial network, thereby achieving a desired influence, perception, or outcome to one or more actors within the target network. This knowledge may also be used to identify key individuals, relationships, and organizational practices. Subsequently, such analysis may lead to the identification of exploitable weaknesses to either eliminate the network as a whole, cause it to become operationally ineffective, or influence it to directly or indirectly support National Security Strategy.

15. SUBJECT TERMS
Layered Networks, Social Network Analysis, Terrorist Networks, Network Flow, Generalized Network Flow, Key Player Problem, Reach-Based Assessment of Position, Decision Analysis, Gains and Losses of Influence

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
			UU	361	Dr. Richard F. Deckro, Professor of Operations Research (ENS)
REPORT	**ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	**U**	**U**			(937) 255-6565, ext 4325; e-mail: Richard.Deckro@afit.edu